REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for re the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jeffers Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

AFRL-SR-AR-TR-02-

0253

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE
	30 Jun 2002	Final

4. TITLE AND SUBTITLE
A frontal attack on limiting defects in nitrides: theory and computations

5. FUNDING NUMBERS
F49620-00-C-0019

6. AUTHOR(S)
Chris G. Van de Walle and John E. Northrup

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
Dr. Gerald L. Witt, Program Manager, Physics & Electronics
AFOSR/NE
801 North Randolph Street, Room 732
Arlington, VA 22203

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Distribution Statement A. Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

Defects are limiting the performance of nitride semiconductor devices: they restrict doping levels, lead to device degradation, and interfere with achieving undoped or semi-insulating material. We have performed a comprehensive theoretical study aimed at facilitating the identification of defects and impurities, and at devising strategies for controlling their formation. Calculations were performed for isolated point defects, defect complexes, and impurities, in a bulk environment, as well as in the vicinity of extended defects and on the GaN surface. In addition to GaN, we performed calculations for InN and AlN, aimed at addressing InGaN and AlGaN alloys that are an essential constituent of device structures. Throughout, the computational results have been used for comparisons with and interpretations of experimental results. Significant results include the identification of a novel structure of Mg-H complexes; a critical assessment of the prospects of codoping; comprehensive results for diffusion of point defects and of hydrogen; a new model for the structure of screw dislocations; detailed values for the effects of impurities on the lattice parameters; the finding that hydrogen acts as a source of *n*-type conductivity in InN; and a comprehensive thermodynamic description of GaN surfaces in the presence of hydrogen.

14. SUBJECT TERMS
Nitride semiconductors; first-principles calculations; defects

15. NUMBER OF PAGES
40

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

20020903 042

A frontal attack on limiting defects in nitrides: theory and computations

Final Report

Period covered: May 1, 2000 – June 30, 2002

Prepared by

Chris G. Van de Walle and John E. Northrup
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

vandewalle@parc.com

For the

Air Force Office of Scientific Research
Contract # F49620-00-C-0019
Program Manager: Dr. Gerald L. Witt

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

Abstract

Defects are limiting the performance of nitride semiconductor devices: they restrict doping levels, lead to device degradation, and interfere with achieving undoped or semi-insulating material. We have performed a comprehensive theoretical study aimed at facilitating the identification of defects and impurities, and at devising strategies for controlling their formation. Calculations were performed for isolated point defects, defect complexes, and impurities, in a bulk environment, as well as in the vicinity of extended defects and on the GaN surface. In addition to GaN, we performed calculations for InN and AlN, aimed at addressing InGaN and AlGaN alloys that are an essential constituent of device structures. Throughout, the computational results have been used for comparisons with and interpretations of experimental results. Significant results include the identification of a novel structure of Mg-H complexes; a critical assessment of the prospects of codoping; comprehensive results for diffusion of point defects and of hydrogen; a new model for the structure of screw dislocations; detailed values for the effects of impurities on the lattice parameters; the finding that hydrogen acts as a source of *n*-type conductivity in InN; and a comprehensive thermodynamic description of GaN surfaces in the presence of hydrogen.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

Objectives

The purpose of this program was to significantly enhance our understanding of native point defects and impurities in III-nitride semiconductors, and the role they play in materials characteristics and device performance. This enhanced understanding will provide the foundation for increased control of impurity and defect incorporation. Through close interaction with experiment, emphasis was placed on those defects that most crucially affect device performance. The objectives stated here are the same as in the original contract.

Computational codes

Our main tool in performing our studies has been a first-principles calculational approach, based on density-functional theory [R1] and *ab initio* pseudopotentials. We use a code that was developed at the Fritz-Haber-Institut in Berlin; our interactions and collaborations with researchers at that institute have provided us with access to state-of-the-art versions of the computational techniques. Further details of the computational approach can be found in Refs. [R2-R5]. The calculations were carried out on IBM RS-6000 workstations and on a Cray T3E available through AFOSR high-performance computing services. For the latter, a highly optimized and parallelized version of the code was used.

We were the first to point out that the *d* electrons in GaN can have a significant effect on the properties of defects or impurities [R6], and we consistently take these effects into account, either through the so-called nonlinear core correction [R7] or by explicit inclusion of the *d* electrons as valence electrons. We also systematically investigate both the wurtzite and zinc-blende phases of the nitride semiconductors.

All calculations were carried out at the theoretical lattice constant [bulk wurtzite GaN: a_{th} = 3.089 Å, compared with a_{exp} = 3.19 Å]. We also used the ideal *c/a* ratio of $\sqrt{8/3}$, which is very close to the calculated *c/a* ratio of 1.633 (exp.: 1.627). We used 32-atom and 96-atom supercells, in the wurtzite structure. Brillouin-zone integrations were

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

carried out with a regular spaced mesh of points in the reciprocal unit cell according to the Monkhorst-Pack method and reduced by symmetry to a set of irreducible **k** points.

Accomplishments

We present the results obtained during the program in the same order, and under the same headings, as listed in the original proposal, to emphasize that all work was carried out as promised. In addition, we will list several topics that arose during the course of the program. For those topics that were covered in previous Progress Reports we only include a brief summary.

1. Formation of complexes

1.a. Complexes between hydrogen and dopant impurities

These studies were described in detail in section "*Vibrational frequencies for hydrogen-related complexes in GaN*" in our Progress Report for the period August 1, 2000 – July 31, 2001, and in publications [P1][1] and [P14]. Summary:

Hydrogen readily incorporates in GaN, particularly in *p*-type material, where it acts as a donor and compensates the electrical activity of the acceptor. Magnesium-hydrogen complexes form during cooldown after MOCVD (metal-organic chemical vapor deposition), and the resulting deactivation of the acceptors explains the need for a post-growth activation treatment [R8]. Knowledge of the atomic configurations in which hydrogen appears before, during, and after the activation process is helpful for optimizing the activation process, and various experimental studies have been devoted to this issue [R9]. Most of the information has been obtained with vibrational spectroscopy, which produces frequencies of the local vibrational modes of the hydrogen-related configurations. These experiments therefore do not provide direct information about the microscopic structure. In order to obtain such information, the experimental results must be combined with computational studies.

[1] We use the notation "[P1]", "[P2]", etc., to cite publications that were produced in the course of the present program; these are included in a separate list at the end of this document. Other references are labeled [R1], [R2], etc.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

We have carried out first-principles calculations for a variety of hydrogen-related configurations, and identified stable and metastable configurations for isolated interstitial H^+ in GaN, as well as for Mg-H complexes. In addition, we have addressed Be-doped material, since we have previously identified Be as a promising alternative acceptor in GaN [R10,P2]. For each of these configurations we have also calculated the local vibrational modes. Comparison with experiment then allows for an unambiguous microscopic identification of the configuration that gives rise to a particular frequency in the vibrational spectrum. *Anharmonic* effects are sizeable for a light impurity such as hydrogen, and they were included for the first time in this type of study. Comprehensive results for all configurations that have been investigated are included in publication [P14]. These results will be very useful for identification of the microscopic structure of species corresponding to vibrational modes observed in infrared or Raman spectroscopy experiments.

An interesting and novel result was found for one of the most important hydrogen configurations, namely the Mg-H complex. This discovery has also revealed some very interesting physics about defects in semiconductors in general. The local vibrational mode (LVM) at 3125 cm^{-1} observed in Mg-doped GaN [R9] was experimentally found to correspond to an N-H bond forming an angle of 130° with the *c* axis of the GaN wurtzite crystal. This large deviation from the expected angle of 109° in the $AB_{N,\perp}$ (antibonding) configuration was very puzzling. The puzzle was resolved by our proposed new configuration, labeled $OA_{//}$ because it originates from an off-axis distortion of a bond-center configuration. The calculated total energy of this configuration is *higher* (by 0.19 eV) than the energy of the $AB_{N,\perp}$ configuration. However, this configuration is stabilized by a large *entropy* contribution to the free energy, caused by a set of low-energy excitations due to the rotational motion of hydrogen around the bond axis.

These results address an important aspect of the technologically important Mg-H interactions in GaN, but they also highlight the importance of taking finite-temperature *free-energy* effects into account when assessing the relative stability of various possible configurations of a defect, impurity, or complex.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

1.b. Complexes between impurities and native defects

These studies were described in detail in section "*Investigations of doping-limiting mechanisms in Mg-doped GaN*" in our Progress Report for the period August 1, 2000 – July 31, 2001, and in publication [P3]. Summary:

Magnesium is almost universally used for *p*-type doping of GaN. However, hole concentrations that can be achieved in Mg-doped GaN films are limited. For example, at 300 K, about 1% of the Mg acceptors are ionized, with a highest reported hole concentration of ~10^{18}cm^{-3} for a Mg concentration of ~10^{20}cm^{-3} [R11]. Increasing the Mg concentration further during MOCVD growth causes the hole carrier concentration to *decrease* and surface roughness to develop. In earlier work [R12], we suggested that the limited *solubility* of Mg in GaN is an important factor. Beyond a certain concentration, Mg is unlikely to incorporate. In fact, precipitates of Mg_3N_2 (which is the solubility-limiting phase) may form, which could contribute to the increased surface roughness. Several experimental studies have shown, however, that deep levels are formed at higher Mg concentration, indicative of compensation. The main compensating center that we have previously addressed is the nitrogen vacancy [R13]. Nitrogen vacancies can be the source of the boue luminescence that is frequently observed in highly doped *p*-type GaN.

The work on this topic carried out during the program was prompted by an experimental study at PARC of films grown by MOCVD with various Mg concentrations. The structural investigations showed that *inversion domains* occurred in GaN films when the bulk concentration of Mg was in excess of 10^{20} cm^{-3}. The inversion-domain formation was accompanied by an *increase* in the Mg incorporation and a *decrease* in the hole concentration.

The observed *reduction* of the hole concentration that occurs when the total Mg concentration [Mg] exceeds 10^{20} suggests that some of the Mg incorporates in an electrically inactive form and/or gives rise to compensating donors at high Mg concentrations. Most of the Mg is incorporated during MOCVD growth in the form of a neutral Mg-H complex, and these Mg atoms act as acceptors only after activation [R8]. In order to account for a reduction in the hole concentration we propose another type of Mg related

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

complex that would act as a *donor* after activation. We focused on one complex to illustrate the concept: a hydrogenated nitrogen vacancy in which two of the surrounding Ga atoms have been replaced by Mg. We refer to this complex as the Mg_2-V_N-H complex. Such a complex would be a *neutral* defect during growth but would become a *donor* once the hydrogen was removed by the activation process.

Our analysis indicates that for high Mg concentrations the concentration of the Mg_2-V_N-H complex becomes comparable to or larger than that of the Mg-H complex. In this latter case the difference between the number of acceptors and donors will be greatly reduced. This model illustrates that if defect complexes that are stabilized under Mg-rich conditions act as donors following activation, then it is possible to account for the existence of a maximum in the hole concentration as a function of [Mg].

1.c. Complexes between native defects

We have investigated instances of complex formation between native point defects, but did not identify any that would be stable enough to play an important role. Some of the configurations that were examined in the course of our studies of point-defect diffusion can be regarded as complexes; for instance, during diffusion of the nitrogen vacancy, the saddle point configuration can be envisioned as a complex between a nitrogen interstitial and two nitrogen vacancies. However, this configuration is not stable (i.e., it is not a local minimum), and only plays a role during diffusion, as discussed in the next section.

1.d. Complexes between impurities; codoping

Codoping (i.e., the incorporation of donors along with acceptors) has been proposed as an effective way of increasing hole concentrations in p-type GaN. Yamamoto and Katayama-Yoshida [R14] have proposed that complexes consisting of two acceptors and one donor (e.g., Mg-O-Mg) would be effective in enhancing the doping efficiency, and would provide an explanation for the experimental observations by Brandt *et al.* [R15] (codoping of cubic GaN with Be and O) and by Korotkov *et al.* [R16] (codoping with Mg and O). Yamamoto and Katayama-Yoshida [R14] performed first-principles calculations, but their arguments were based mainly on trends in the electrostatic (Madelung) energy.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

We have performed comprehensive investigations of Be-O-Be complexes [P2] and of Mg-O-Mg complexes. We found that the formation energy of these complexes is *not* lower than that of the isolated acceptors (at least if equilibrium with the proper solubility-limiting phases is taken into account). Formation of acceptor-O complexes (which are electrically neutral) is energetically quite favorable, but attaching a second acceptor impurity to such a complex is only marginally favored. Finally, even if Be-O-Be or Mg-O-Mg complexes would form, there is no indication that their ionization energy is lower than that of isolated acceptors (within the error bars of our calculations). We therefore doubt that formation of Be-O-Be or Mg-O-Mg complexes is a viable approach to increasing *p*-type doping of GaN, or provides an explanation for the experimentally observed *p*-type conductivity [R15, R16] in codoped samples.

2. Migration of defects and impurities

Knowledge of the diffusion properties of native point defects is important to assess the likelihood of their incorporation during growth and processing; in addition, it forms the basis for understanding impurity diffusion, which is nearly always mediated by native defects. Diffusion also plays an important role in device degradation.

Information about diffusion in nitrides is currently very limited. Only a few experimental studies have been performed, resulting in mostly qualitative information. Furthermore, the results may also reflect phenomena induced by the presence of a large concentration of extended defects, and therefore this information is probably of limited applicability to bulk GaN. We have performed a comprehensive computational study tracking the diffusion paths of various relevant point defects. The main results are reported here, and are documented in a manuscript that has been prepared for publication [P17].

2.a. Formation energies of native point defects

As a starting point, we re-examined the formation energy of all native point defects in their ground-state configurations for all relevant charge states. We performed similar investigations in 1994 [R5], but at that time computational limitations restricted the size of

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

the supercells to 32 or maximum 72 atoms. Expanded computer power now allows us to perform calculations in 96-atom supercells for the wurtzite structure, and we have updated the results for vacancies and self-interstitials; antisites are high in energy and are unlikely to affect the materials properties. The formation energies for all native point defects obtained in this fashion are shown in Fig. 1. Gratifyingly, the results are very similar to those reported in Ref. [R5]. For quantitative accuracy in our diffusion studies, however, it was important to consistently use results obtained in same-size supercells when comparing ground-state energies with saddle-point energies.

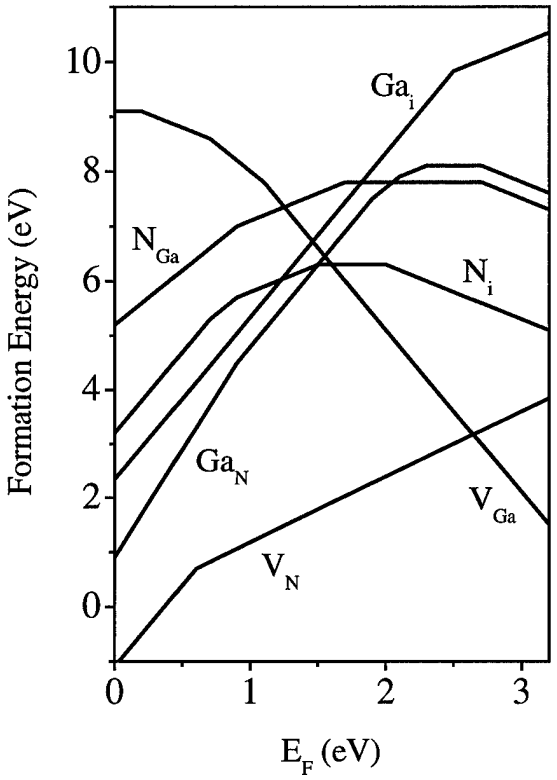


Figure 1: Formation energies as a function of Fermi level for native point defects in GaN. Ga-rich conditions are assumed. The zero of Fermi level corresponds to the top of the valence band. Only segments corresponding to the lowest-energy charge states are shown. The slope of these segments indicates the charge state. Kinks in the curves indicate transitions between different charge states.

For our investigations of point-defect diffusion, we focused on self-interstitials and vacancies. Antisites are high in energy (see Fig. 1) and we have no reason to believe they play an important role under either equilibrium or non-equilibrium conditions.

2.b. Gallium interstitials (Ga_i)

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

Detailed experimental data are available for the gallium interstitial. Using electron paramagnetic resonance, Chow *et al.* [R17] studied point defects created by electron irradiation at low temperature. They were able to unambiguously identify the Ga interstitial, and their annealing studies indicated that this interstitial becomes mobile at temperatures slightly below room temperature. An activation energy of 0.7 eV was estimated.

Our calculations indicate that in wurtzite GaN, the gallium interstitial favors a site near the octahedral interstitial site, at the center of the hexagonal channel. We call this site *O'* to distinguish it from the ideal octahedral interstitial *O* site. As shown in Fig. 1, the stable charge states are 3+ and 1+, meaning that Ga_i always acts as a donor. Its formation energy is lowest (but still higher than that of the nitrogen vacancy) for Fermi-level positions near the valence-band maximum (VBM), i.e., under *p*-type conditions. The +2 state is not thermodynamically stable, and the 1+ charge state is stable at higher Fermi levels, where the high formation energy renders its formation unlikely under equilibrium conditions. However, Ga interstitials can be induced by non-equilibrium processes, such as in the irradiation experiments of Ref. [R17].

The high formation energy of the Ga interstitial is caused by the large size of the Ga atom, which does not "fit" inside the interstitial space available in the GaN lattice with its small lattice parameters. One might therefore think that the Ga interstitial will also have trouble moving through this lattice, i.e., that the migration barrier will be high. This is not the case. In fact, the migration barrier is the energy difference between the saddle-point and the ground-state configurations; if *both* are high in energy, their difference can be modest, resulting in a moderate barrier.

We investigated migration paths parallel to the *c* axis ($\| c$) as well as in directions perpendicular to the *c* axis ($\perp c$). In these investigations we were aided by our previous work on diffusion of beryllium interstitials in GaN (also carried out during this program, see Ref. [P2]). First we tracked the migration path along the hexagonal channel, resulting in a barrier height of ~3.0 eV. Further investigations showed, however, that another migration path exists with a much lower barrier. In the case of the Be interstitial [P2], the migration proceeded by the interstitial moving from its stable configuration at *O'* through

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

a saddle point at the tetrahedral interstitial site (*T*) to an *O'* site within the same basal plane in an adjacent hexagonal channel. This whole process occurred with a much lower migration barrier (1.2 eV) than migration through the hexagonal channel (which cost about 3 eV for Be_i). A similar situation occurs for Ga_i, but now the diffusing atom is the same element as one of the host atoms, and therefore diffusion can proceed via the so-called interstitialcy mechanism.

Figure 2 shows a plot of the potential energy along the migration path. This curve indicates that a local minimum occurs along the path; we call this the *A* site. At the *A* site the interstitial forms a split-interstitial configuration with a Ga host atom. When the interstitial is moved further towards *T,* it actually replaces the Ga host atom; i.e., it knocks the Ga host atom out of its lattice site, forcing it to become the new interstitial, which will move towards an *O'* site in an adjacent hexagonal channel. This new *O'* site is located in a basal plane different from the site that was the starting point; i.e., this migration path can apply to diffusion both parallel to *c* and perpendicular to *c* – unlike the case of the Be interstitial, where a clear anisotropy in the diffusion occurred. As shown in Fig. 2, the barrier along the migration path is less than 1.0 eV.

Our most detailed investigations were carried out for the Ga interstitial in the 3+ charge state. However, our studies show a very similar migration path, with possibly slightly lower barriers, for the +2 and +1 charge states. In the experiments of Ref. [R17], *n*-type material was used, and therefore the relevant charge state was likely to be +1. Our results, which show that migration of the Ga interstitial can proceed with a barrier of less than 1 eV, are in good agreement with the experimental observations of Ref. [R17]. Finally, we note that the transition level between the +3 and +1 charge states of the Ga interstitial occurs at 2.5 eV above the valence band in our calculations. This position of the defect level is consistent with the PL emission around 0.9 eV that has been observed in samples containing Ga interstitials [R17].

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

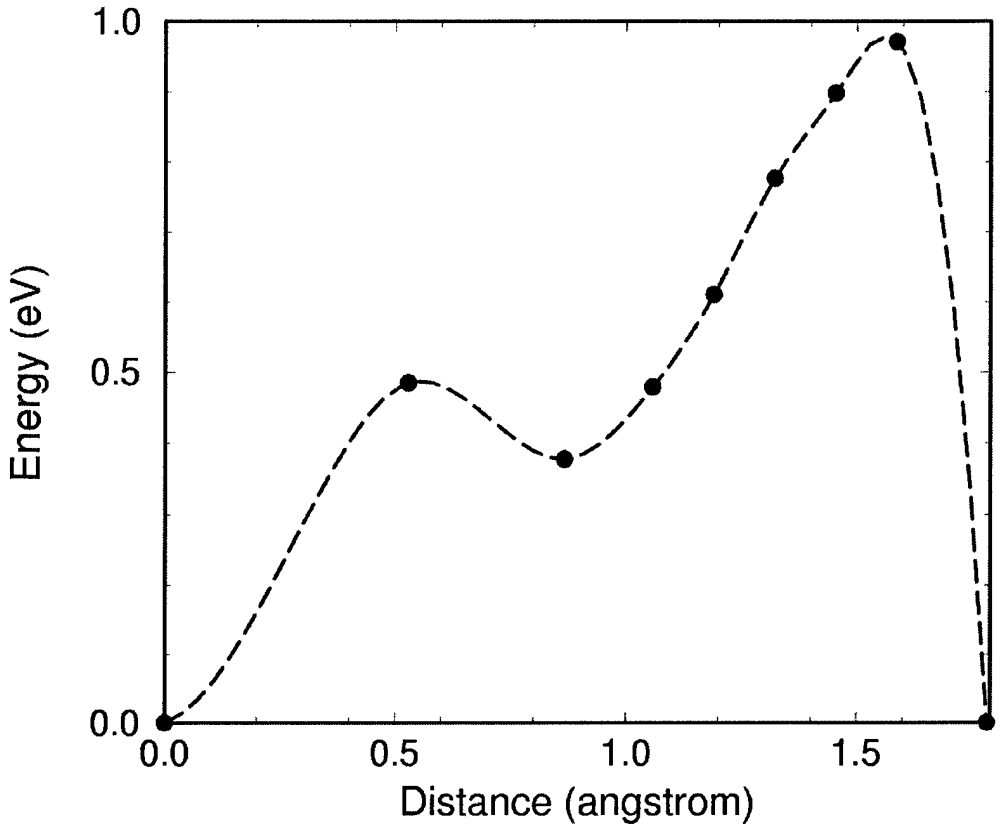


Figure 2: Total energy (referenced to the ground-state configuration) of Ga_i^{3+} as a function of the horizontal distance of the Ga_i atom measured from the center of the hexagonal channel towards the *T* site. By the time the Ga_i has reached the *T* site, it has moved through a triangle of N atoms and replaced a Ga host atom, pushing the original Ga host atom off its lattice site to become the new Ga_i.

2.c. Nitrogen interstitials (N_i)

For nitrogen interstitials, the ground state consists of a split interstitial in which the N_i forms a short bond with one of the host N atoms, sharing its lattice site. This configuration is energetically preferred, by a large margin, over other interstitial positions such as the octahedral (*O*) or tetrahedral (*T*) interstitial sites, due to the large binding energy of the N_2 molecule. N_i can occur in various charge states, as shown in Fig. 1.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

The details of the atomic structure vary with the charge state of the defect, with the N-N bond distance varying from 1.45 Å in the 1– charge state to 1.12 Å in the 3+ charge, but the split-interstitial configuration is common to all charge states. Note that the N-N bond distances are comparable to the bond distance of N_2 (1.0975 Å). Since the host atom shares its lattice site with the interstitial, it is not clear which of the two atoms is the interstitial N; both should be regarded on equal footing. Nitrogen interstitials have fairly high formation energies for all Fermi level positions (Fig. 1). They are thus unlikely to occur in thermal equilibrium. However, their formation may be induced under non-equilibrium conditions, for instance under irradiation [R18,R19].

Nitrogen interstitials diffuse via an interstitialcy mechanism as illustrated in Fig. 3. Note that the atomic positions in Fig. 2 do not reflect the actual relaxations; the figure merely serves as a schematic for illustrating the diffusion process. As expected, the saddle point occurs near the position where the traveling N atom is equidistant between two nominal nitrogen lattice sites. Our extensive investigations of the migration paths resulted in the following migration barriers for the various charge states: 1.6 eV for 1–, 2.4 eV for 0, 2.1 eV for1+, 2.5 eV for 2+, and 1.4 eV for 3+. Our calculated migration barriers indicate that nitrogen interstitials could become mobile at temperatures around 200 °C, at least for the – and for the 3+ charge states (which are probably most likely to occur, under *n*-type, resp. *p*-type conditions).

In addition to the diffusion process discussed above, N_i could also move in a direction perpendicular to [0001] ($\perp c$). This involves one of the N atoms in the split-interstitial configuration breaking away from its pair and traveling in the basal plane to bond with another N in the same plane. Our calculations indicated that this barrier is higher than the barrier for the process illustrated in Fig. 3. Since the process of Fig. 3 can result in motion both parallel to *c* and perpendicular to *c*, diffusion along both directions will be governed by the same barrier; i.e., no anisotropy will occur.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

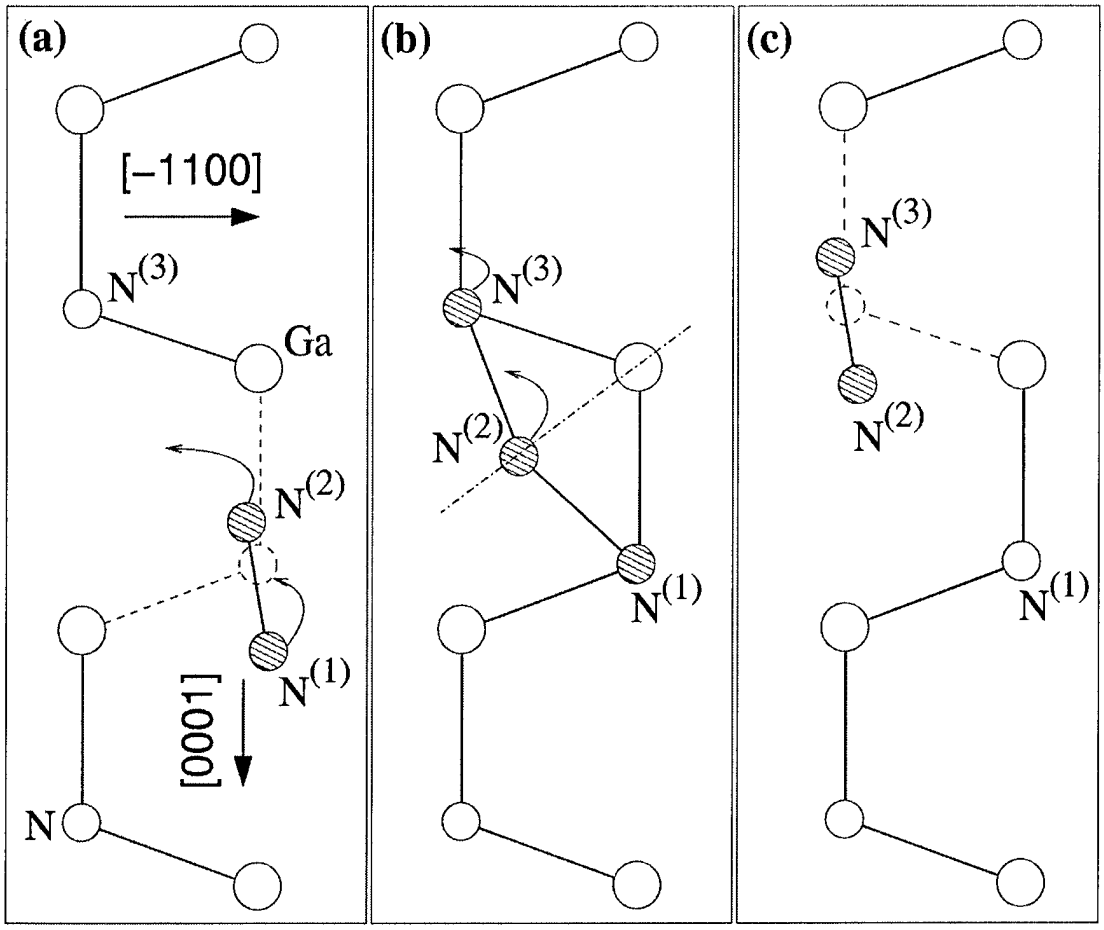


Figure 3: Schematic illustration of the migration of interstitial nitrogen via an interstitialcy mechanism. (a) Ground state. (b) Saddle-point configuration. (c) New ground state, after one migration step. Large circles represent Ga atoms, medium circles N atoms, and shaded medium circles highlight N atoms involved in the migration. Dashed circles indicate ideal atomic position of the lattice N, dashed lines bonds in the ideal lattice. The numbers in square brackets identify the migrating atoms. The center of mass of the two closest nitrogen atoms is also indicated.

We have also performed investigations of a complex consisting of *two* nitrogen interstitials ($2N_i$). Together with a N host atom to which they bind, these nitrogens can form a *trimer*. Alternatively, the complex can be regarded as an interstitial N_2 molecule. For Fermi-level positions in a range around midgap, this new complex has a lower for-

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

mation energy than two separate N_i. Various distinct configurations of the $2N_i$ complex were found to have very similar energies. In addition, a complex composed of two N_i in split-interstitial configurations on adjacent N lattice sites was also found to have lower formation energy than the isolated N_i. The complexity of investigating migration of a complex with multiple constituents has precluded us from obtaining a firm value for the migration barrier. However, the similarity in formation energies among different configurations may indicate that this complex could migrate with rather low migration barriers.

Such a complex could then be a candidate to explain the migration phenomena that have been observed to occur at very low temperatures. For instance, the low-temperature irradiation studies by Chow *et al.* [R17] indicated that the *L*1 defect is not present right after irradiation, but starts forming by annealing at temperatures as low as 135 K. Motion at this temperature requires a migration barrier as low as 0.4 eV. In other experiments, Zhou and Manasreh [R19] reported that the optical absorption associated with nitrogen vacancies in AlGaN annealed out with an activation energy around 0.4-0.5 eV. As we will see below, the migration barrier of the nitrogen vacancy itself is almost an order of magnitude larger than this value – but nitrogen vacancies could of course also be annihilated due to recombination with mobile nitrogen interstitials. As reported above, the isolated nitrogen interstitials have migration barriers of at least 1.4 eV. Our investigations of complexes consisting of two nitrogen interstitials suggest that they could be responsible for the moving species with a barrier around 0.4 eV.

2.d. Nitrogen vacancies (V_N)

Nitrogen vacancies have low enough energies to occur under thermal equilibrium conditions in *p*-type GaN. The stable charge states for V_N in GaN are 1+ and 3+. The 2+ charge state is never stable; this is characteristic of a negative-*U* impurity, and is usually associated with a strikingly large lattice relaxation of one of the charge states. Indeed, in the ground state, the neighboring atoms V_N^+ relax outward by (on average) 3.7 % (expressed as a fraction of the Ga-N bond). But in the ground state of V_N^{3+}, the neighboring

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

atoms relax outward by 14.6 % on average. The transition level $\varepsilon\{3+/+\}$ between the 3+ and + charge states was found at 0.59 eV above the valence-band maximum (VBM). This means the 3+ charge state is stable under *p*-type conditions, and its low formation energy indicates that nitrogen vacancies can be a serious source of compensation in *p*-type GaN. Note that the formation energy of V_N^+ increases with the Fermi level (see Fig. 1), so that nitrogen vacancies are unlikely to occur in *n*-type GaN – a conclusion that seems to be well accepted now but was quite controversial when we first reported it [R5].

We obtained the migration barrier of the nitrogen vacancy by calculating the energy needed to move a nitrogen atom from its nominal lattice adjacent to the vacancy along a path towards the vacancy, leaving a vacancy behind at the original lattice site. The path is illustrated in Fig. 4.

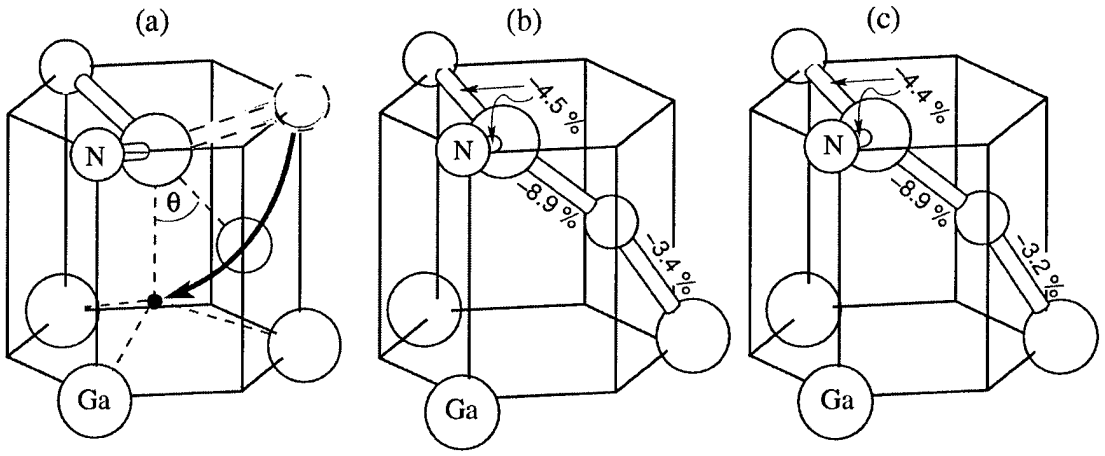


Figure 4: Ball-and-stick model of atomic positions describing the diffusion process of the nitrogen vacancy. (a) Illustration of the migration path. (b) The saddle-point configuration of V_N^{3+}. (c) The saddle-point configuration of V_N^+. Large spheres represent Ga atoms, medium spheres N atoms, and the black dot shows the ideal lattice site of the vacancy. The percentages denote the changes from the bulk Ga-N bond length.

We define the coordinate along the path as the angle θ of the bond between the migrating N atom and the ideal lattice site of the Ga atom at the pivot point, measured

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

from the [0001] direction. The saddle point along the migration path occurs at an angle θ of about 50°. For the nitrogen vacancy (V_N), we find a high migration barrier for V_N^+ (>4 eV), suggesting V_N^+ will only be mobile near the growth temperature. V_N^{3+}, on the other hand, which can form in *p*-type material, exhibits a much lower migration barrier of 2.6 eV. We propose the following explanation for this large difference: for V_N^+, the saddle-point configuration can be regarded as a complex consisting of two nitrogen vacancies and one nitrogen interstitial: $2V_N^+ + N_i^-$, where the charges have been chosen such that the total charge of the complex (1+) is maintained. Similar arguments for the saddle point of V_N^{3+} suggest a complex consisting of $2V_N^+ + N_i^+$. The energy of the $2V_N + N_i$ complex will differ mainly due to the different charge on N_i (assuming no change in binding energy of the complex). If we now assume that the Fermi level is 0.5 eV above the VBM (a typical value if we want V_N^{3+} to be stable) then the energy difference between $2V_N^+ + N_i^-$ and $2V_N^+ + N_i^+$ is 2.5 eV. This difference is consistent with the large difference in migration barriers between V_N^+ and V_N^{3+}.

The migration path described above can account for motion along the *c* axis (∥ *c*) as well as in directions perpendicular to the *c* axis (⊥*c*). We have also studied a migration path that would account exclusively for in-plane (⊥*c*) diffusion, by moving a nitrogen atom from the same basal plane as the vacancy towards the vacancy. Note that this path would be relevant only if it results in a lower barrier than the barriers calculated above. This is because diffusion ⊥*c* can also be accomplished via the diffusion process described above, which creates lateral as well as vertical displacement. Our calculated migration barriers for the in-plane motion were *higher* than the ones calculated above for both charge states. Therefore, migration of V_N^+ and V_N^{3+} will be governed by the same barriers for diffusion ∥ *c* and ⊥*c*. No anisotropy should be observed.

2.e. Gallium vacancies (V_{Ga})

Our investigations of gallium vacancies were performed using an approach similar to that used for nitrogen vacancies, i.e., a neighboring Ga atom was moved towards the vacancy

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

and the potential energy along that path was evaluated. A migration barrier of 2.0 eV was obtained for V_{Ga}^{3-}, the most relevant charge state.

Our results for gallium vacancies are in good agreement with the positron annihilation experiments of the group of Saarinen [R20,R21]. In Ref. [R20] they showed that concentrations of gallium vacancies are higher in *n*-type material than in *p*-type or semi-insulating material, consistent with our calculated formation energies (Fig. 1). They also showed that gallium vacancies have higher concentrations in oxygen-rich material, consistent with our prediction [R22] that a V_{Ga}-O complex can form with a binding energy of 1.8 eV.

Note that our theoretical work indicates that the gallium vacancy could give rise to a variety of transitions in the yellow and green region of the spectrum. The transition level between the 3- and 2- charge states of the isolated vacancy occurs at 1.1 eV above the VBM ([R22] and Fig. 1]), which formed the basis for our prediction that gallium vacancies are responsible for the "yellow luminescence". We have also pointed out, however, that the 2-/1- transition level occurs at only slightly lower energy (which would push the emission towards the green). Complex formation with oxygen also pushes the level down (i.e., towards green emission). These results may well explain the observations of green and yellow luminescence in freestanding GaN templates by Reshchikov *et al.* [R23].

Saarinen *et al.* also applied positron annihilation spectroscopy to study thermal annealing of gallium vacancies [R21]. Vacancies induced by electron irradiation were found to anneal out in long-range migration processes at 500–600 K with an estimated migration energy of 1.5 eV, in reasonable agreement with our calculated migration barrier of 2.0 eV. Native Ga vacancies in as-grown GaN, on the other hand, survived up to much higher temperatures (1300–1500 K), leading to the conclusion that they are stabilized by forming complexes with oxygen impurities. The estimated binding energy of 2.2 eV of such complexes is in agreement with our theoretical predictions [R22].

As an additional prediction, we comment on self-diffusion. The activation energy for self-diffusion is the sum of the formation energy of the defect and the migration en-

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

ergy. Under *n*-type conditions, the formation energy of V_{Ga}^{3-} is in the range of 1 to 2 eV. Adding the migration energy, the activation energy for self-diffusion would be between 3 and 4 eV. An activation energy of 3 eV would imply that self-diffusion due to gallium vacancies could occur at temperatures around 800 °C. This is certainly consistent with the notion that gallium vacancies play a role in interdiffusion of InGaN quantum wells, as discussed in Ref. [23] (see below).

2.e. Hydrogen

Hydrogen is a common unintentional impurity in GaN. Knowledge of its diffusivity is important, for instance for optimizing the thermal activation process that is necessary to render acceptor-doped GaN grown by MOCVD *p*-type conductive. We performed a detailed study of the stability and diffusivity of monatomic (H^+ and H^-) and molecular (H_2) hydrogen in wurtzite GaN, as documented in Ref. [P15]. Diffusion barriers in the direction parallel to [0001] ($\| c$) and perpendicular to [0001] ($\perp c$) were calculated. For H^+ the diffusion barriers are low and almost isotropic at 0.9 eV, slightly higher than in the zinc-blende phase. For H^- the diffusion barriers are significantly lower than in zinc-blende GaN, with values of 2.0 eV ($\| c$) and 2.2 eV ($\perp c$). The diffusion barriers for H_2 are relatively high: 2.0 eV ($\| c$) and 2.3 eV ($\perp c$), and we propose that diffusion of H_2 is more likely to proceed by dissociation followed by diffusion of monatomic H^+.

3. Relation between point defects and extended defects

3.a. Threading dislocations in GaN: the screw dislocation

These studies were described in detail in our Progress Report for the period August 1, 2000 – July 31, 2001, and in publications [P7] and [P12]. Summary:

Dislocations are abundant and problematic in GaN epitaxial films and devices. They are sources of non-radiative recombination and pathways for current flow under reverse bias conditions. And although it is clear that the electrical activity depends on the atomic structure of the dislocation core, little is known about the core structure. Consequently, there is a need for both experimental and theoretical study of dislocations. This

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

need was addressed in our theoretical study of screw dislocations in GaN. Possible models for the pure screw dislocation were developed and density-functional-theory calculations of the relative stability of these models were performed. The two standard models for the screw dislocation are the full core model and the hollow core model. In the full core model the stoichiometry is the same as that of bulk GaN, however, the atoms near the core are very highly strained. The hollow core model is obtained by removing all of the highly strained atoms from the core. This results in the presence of Ga and N dangling bonds on the walls of the hollow core. Our calculations predict that neither of these models is an equilibrium structure. In Ga-rich growth conditions the equilibrium structure is a core containing only Ga atoms, i.e. all the N atoms are missing from the core region. The Ga atoms in the core form a pair of helical chains. This new model provides a natural explanation for Z-contrast TEM images that appear to show a filled core [R25]. As one moves back towards N-rich conditions it is energetically favorable to add N back into the core, and in the N-rich limit the lowest energy structure contains both N and Ga atoms. Electronic structure calculations were performed and it was found that the dislocation would be electrically active in both Ga-rich and N-rich conditions. The electrical activity can be traced to the existence of dangling bonds that are present on Ga atoms. These results can be used to interpret the recent experimental results by Hsu *et al.* [R26], who found that screw dislocations are sources of non-radiative recombination and can give rise to large leakage currents.

3.b. Interaction of hydrogen with dislocations

An important question is whether the hydrogen that is present in MOCVD growth of GaN can interact strongly with dislocations and change the atomic structure and electrical activity. We therefore initiated an investigation into this possibility. Specifically we investigated a model in which all of the N and Ga dangling bonds that would be present in the hollow core model are saturated by H atoms. This structure contains no dangling bonds and hence this type of dislocation would be electrically inactive. The relative stability of this structure with respect to the other possible models (without hydrogen) was

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

determined as a function of the hydrogen chemical potential. The hydrogen chemical potential under MOCVD growth conditions can be obtained as a function of the gas phase pressure and the temperature. In this way it was determined that in Ga-rich conditions the stability of the H-passivated dislocation requires the temperature to be below about 800 K and in N-rich conditions the temperature must be below 1200 K. In both cases it appears that the typical MOCVD temperature of 1350 K is too high to allow the stability of the H-passivated cores. However, hydrogenation at lower temperatures should cause the core structure to change. This work is described in a paper that will be published in Physical Review B [P12].

3.c. Interaction of silicon with screw dislocations

The question of how dopants and other impurities interact with dislocations is an important one, and we performed research to address this question in the case of Si. Silicon is a donor in GaN and one expects Si donors near dislocations to become ionized by transference of an electron to an unfilled dislocation state in the gap. One would then expect the electrostatic interaction to give rise to an attractive force between the ionized Si atom and the dislocation. Calculations were performed for Si replacing three different types of Ga sites near the core of the dislocation. These sites are designated as the bulk site, the wall site, and the helix site. In the bulk site the Si atom is four-fold coordinated. It is bonded to four N atoms – as is a normal Si donor in bulk GaN. In the wall site the Si is bonded to three N atoms and has a dangling bond. In the helix site the Si is bonded to two Ga atoms and to one N atom. The bulk site has the lowest energy and relative to it the energies of the wall site and the helix site are higher by 1.2 and 2.0 eV [P19]. This result indicates that in addition to electrostatic energy considerations – which should favor the wall and helix sites – the local bonding properties of the impurity play an important role in governing the energetics. In this case the preferred site for the Si is determined by the preference for forming as many Si-N bonds as possible.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

These results illustrate the important interactions that can take place between impurities and extended defects. We believe that the interaction of other impurities and point defects with dislocations is an interesting area for future research.

4. Effect of impurities on the lattice parameters

Our results were described in detail in our Progress Reports for the periods May 1, 2000-July 31, 2000 and August 1, 2000 – July 31, 2001; they have been submitted for publication [P13]. Brief summary:

Nitride semiconductors frequently contain large concentrations of impurities, which are introduced either intentionally (in the case of dopants such as Mg or Si) or unintentionally (in the case of contaminants such as oxygen). Thin epitaxial nitride films are also usually strained. The role of impurities as a cause of this strain has been debated; we have performed a first-principles investigation of these effects. The key results are:

4.a. Silicon

Incorporation of silicon has only a minor effect on the lattice parameters. This allows us to conclude that the changes in lattice constant that have been observed as a function of Si concentration [R27] can *not* be attributed to lattice distortions caused by the substitution.

4.b Oxygen

Incorporation of oxygen leads to an expansion of the lattice. For instance, for an oxygen concentration of 10^{20} cm^{-3} we find a lattice expansion $\Delta a/a$=0.0002, which is measurable in X-ray diffraction. This finding impacts the accepted values of the lattice parameters of "pure" bulk GaN. These parameters have usually been obtained from measurements on bulk crystals [R28]. These bulk crystals have been revealed to contain large concentrations of oxygen [R29]. We suggest that the observed difference in lattice constants of bulk crystals between "undoped" GaN and "Mg-doped" GaN [R29] is due to the incorporation of oxygen.

4.c. Magnesium

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

Incorporation of Mg leads to a fairly large lattice expansion. Incorporation of Mg in the large concentrations required to achieve adquate *p*-type conductivity may thus lead to an observable change in the lattice constant.

5. Defects in alloys

All nitride-based devices rely on heterojunctions between GaN and other III-nitride-based alloys, and hence the materials properties of these alloys need to be thoroughly understood. AlGaN layers play a crucial role in barrier layers for electronic devices such as high electron mobility transistors (HEMTs), as well as in lasers, and InGaN layers constitute the quantum wells at the heart of light emitters. Performing explicit calculations for defect configurations is a complex and time-consuming task; indeed, a comprehensive study would involve calculations for a number of different alloy compositions, and for each composition, consideration of a multiplicity of different configurations. For instance, a nitrogen vacancy in InGaN could be surrounded by 4 Ga, or 3 Ga and 1 In, or 2 Ga and 2 In, etc. – and for several of these possibilities different configurational arrangements might be possible. We have therefore restricted ourselves (with one exception) to the more limited task of performing defect calculations in the binary compounds (InN and AlN, in addition to GaN), and drawing conclusions about alloys based on linear interpolation.

The one exception we are alluding to is a series of calculations for gallium vacancies in InGaN alloys. As reported in Ref. [R24], gallium vacancies are likely to play a role in the observed phase separation in InGaN multiple quantum wells annealed at high nitrogen pressures. Such phase separation is caused by diffusion of the constituents, and is a very serious technological issue for InGaN quantum wells in light-emitting devices. Our calculations for gallium vacancies in InGaN alloys (for one particular alloy composition, namely 6.25%) indicated that the formation energy of V_{Ga} is lowered when the vacancy is created in the vicinity of an In atom. This results in enhanced formation of such vacancies in InGaN layers, which may contribute to the observed interdiffusion [R24]. This lowering of the formation energy is an alloy-specific effect that was not expected on

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

the basis of the formation energies of cation-site vacancies in GaN (Fig. 1) and InN (Ref. [R30]).

Overall, however, interpolation between the binary compounds can provide a reasonable idea of the behavior of point defect energies in alloys. For InGaN, the point-defect energies in InN were reported in Ref. [R30]. For AlGaN, we performed a comprehensive investigation of point defects in AlN in the course of the present program (see Ref. [P6]. The main results are that (1) V_N^{3+} has the lowest formation energy in *p*-type material, and V_{Al}^{3-} has the lowest formation energies in *n*-type material; the dominating compensating centers are therefore similar to the corresponding ones in GaN; and (2) the formation energy of the nitrogen vacancy is high in *n*-type AlN, indicating that this defect is not expected to occur in high concentrations.

The electronic structure of the nitrogen vacancy is qualitatively different in AlN compared to GaN: in GaN the higher lying defect-induced level is a resonance in the conduction band, but in (wurtzite) AlN it lies well below the conduction-band edge. The nitrogen vacancy is thus a shallow donor in GaN, but a deep donor in wurtzite AlN. Interpolation of these results between GaN and AlN leads to a position of the V_N-induced defect level in agreement with the measurements in the group of M. O. Manasreh [R18, R19].

While we believe our present study has produced important conclusions, additional work will clearly be required in order to fully understand and control defect formation and doping in alloys. In particular, AlGaN layers with increasingly high Al content are being pursued for electronic and optoelectronic (UV emitting) devices. Such alloys are known to be very difficult to dope (both *p*-type and *n*-type), and new types of detrimental radiative as well as nonradiative recombination channels appear as the band gap increases. Additional investigations of dopants, unintentional impurities, and isoelectronic centers, and of the interaction of point defects with all of the above, would be highly beneficial.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

6. Defect creation driven by polarization fields

Polarization fields play an important in the nitrides. Even in the absence of strain (which leads to piezoelectric fields), spontaneous polarization is present. In macroscopic samples the resulting electric fields are screened, but on microscopic length scales the fields could still provide a driving force for defect creation. Large fields also occur in heterostructures and quantum wells. We originally planned to investigate how the formation energy of native point defects is affected by the presence of these fields. This task turned out to be more difficult than anticipated due to two problems: (1) the technical issues associated with incorporating fields in the first-principles calculations; and (2) the uncertainty that still existed regarding the magnitude of polarization fields, and the extent to which they are screened.

In the course of the program we were actually able to contribute to progress in the latter area, in conjunction with electroabsorption measurements [R31]. The electroabsorption technique allows the exact quantitative analysis of absorption and absorption changes in the InGaN quantum wells and barrier layers. The technique thus allows a precise determination of the strong internal fields that originate from strain-induced polarization and differences in the spontaneous polarization. The measurements were performed on *pin* diodes with the InGaN-layer inside the space-charge region, ensuring that the piezoelectric/internal fields are not influenced by free carriers. Theoretical modeling of the device structures contributed to interpretation of the experimental results. The fields measured on functioning diodes varied between 1.1 MV/cm and 1.4 MV/cm for indium concentrations in the InGaN quantum wells ranging from ~ 7% to 9%. These field strengths are larger than previously extracted from experimental measurements.

7. Role of unintentional impurities

7.a. Boron

Our results for boron were described in detail in our Progress Report for the period May 1, 2000-July 31, 2000. Brief summary:

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

It is often assumed that isoelectronic impurities are electronically inactive. For instance, incorporation of As in GaN has traditionally been assumed to lead to substitution on the N site. In small concentrations, As was assumed to have no effect, while in large concentrations it was assumed to lead to formation of GaAsN alloys. We recently showed [R32] that As may actually also incorporate on the Ga site. The driving force is mainly the large size mismatch that exists between As and N: placing As on a N site creates large strains, whereas As "fits" quite nicely on a Ga site. There is, of course, an "electronic" mismatch for As on a Ga site: since As is two columns to the right of Ga in the periodic table, it acts as a double donor.

It is interesting to contemplate whether similar effects may occur for other isoelectronic impurities. Boron is a particularly important case. It may occur as an unintentional impurity during MBE growth [R33], and there has been some speculation that boron might act as an acceptor in GaN. The driving force for this behavior would again be a preference for the impurity to occupy the site on which the size mismatch is minimized: B would prefer the N site, from this point of view. On this site, it should act as a double acceptor.

Our first-principles calculations show that B overwhelmingly prefers the Ga site, although even on this site its formation energy is rather high. Formation of BGaN alloys will thus be difficult, and may require conditions far from equilibrium. The configuration with B on a N site is much higher in energy, and thus exceedingly unlikely to form. We also find that B_{Ga} is a *deep* acceptor, and would not be useful for *p*-type doping.

7.b. Hydrogen in nitride semiconductors

These studies were described in detail in our Progress Report for the period August 1, 2000 – July 31, 2001, and in publication [P5]. Summary:

Hydrogen is an unintentional impurity that can have a major effect on the properties of nitride semiconductors. Since hydrogen is present in many of the commonly used growth environments for nitrides, its incorporation (if energetically favorable) can hardly be avoided. We previously performed a comprehensive investigation of hydrogen in GaN

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

[R8,R34], and found that H acts as a donor (H^+) in *p*-type GaN, and as an acceptor (H^-) in *n*-type GaN. Hydrogen is thus an *amphoteric* impurity, always acting as a compensating center and counteracting the dominant electrical conductivity of the material.

In the course of the present program we investigated hydrogen in AlN and InN [P5]. Indeed, device structures usually incorporate layers of InGaN and/or AlGaN, and also pure AlN and InN are being pursued in their own right; a thorough understanding of hydrogen in these materials is therefore essential. Our results for AlN show that the behavior of H in AlN is very similar to GaN: H^+ dominates in *p*-type, H^- in *n*-type. Due to the larger band gap of AlN, the solubility of H could be significantly larger than in GaN under both *p*-type and *n*-type conditions.

Our results for H in InN, however, revealed a true surprise: H in InN behaves exclusively as a donor. I.e., it is *not* amphoteric as in GaN and AlN, but actually contributes to the *n*-type conductivity of the material. This donor behavior is due to the fact that H^0 and H^- are always higher in energy than H^+, making H^+ the *only* stable charge state for all positions of the Fermi level.

This result may help explain why, to date, InN samples have typically been found to be highly *n*-type. Combined with our previous results for oxygen and silicon as impurities in InN [R30], we find that that all potential donor impurities in InN have relatively low formation energies, making them easy candidates for unintentional incorporation. Of all the donors, hydrogen actually has the lowest formation energy. After publication of our results [P5], the experimental work of Look *et al.* [R35] confirmed the suggestion that hydrogen may be an important donor in InN.

7.c. Carbon in GaN

Carbon is a common unintentional impurity in GaN. For instance, in MOCVD growth it can be incorporated from the metal-organic sources. Carbon is known to act as an acceptor, but an outstanding question is whether it acts as a deep or as a shallow acceptor. Although we performed a large number of calculations on carbon acceptors during the course of this program, we are as yet unable to answer this question regarding its ioniza-

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

tion energy. The electronic behavior of carbon turns out to be exceedingly sensitive to the details of the first-principles calculations, such as choice of exchange-correlation potential, use of local-density approximation *versus* generalized gradient approximation, use of the nonlinear core correction, etc. Theoretical resolution of the issue of the carbon ionization energy will have to await further computational developments, which we are currently planning.

We did obtain results on another important issue, namely the structure and energetics of the C-H complex. Since both carbon and hydrogen are abundantly present during MOCVD growth, such complexes are likely to be incorporated. We find the lowest energy configuration for hydrogen bonded to the carbon atom (on a nitrogen site, C_N), with H located in an antibonding position ($AB_\perp$). The $AB_\parallel$ configuration (with the C-H bond oriented along the *c* axis) is about 0.3 eV higher in energy.

The calculated binding energy of the C-H complex is 1.2 eV. This is to be compared with a binding energy of 0.7 eV for Mg-H [R34], and 1.8 eV for Be-H [P2]. The comparison would indicate that dissociation of C-H complexes may require slightly higher temperatures than dissociation of Mg-H. However, the temperatures that are commonly applied during annealing of Mg-doped GaN are much higher than strictly needed for breaking the Mg-H bond, indicating that kinetic barriers to neutralization or removal of hydrogen from the layer are playing the dominant role. That would imply that similar temperatures should suffice to remove H from the vicinity of carbon.

8. Role of native defects versus impurities in obtaining semi-insulating material

An important goal of programs aimed at nitride-based electronic devices is to obtain high-quality semi-insulating material. GaN still has a tendency to be unintentionally *n*-type. We have argued for many years that reducing impurity concentrations (oxygen, silicon, ...) is the prime way of reducing *n*-type conductivity. One can wonder, however, whether once impurities are totally under control, point defects would become an issue. Our first-principles formation energies allow us to evaluate defect concentrations as a function of doping level and growth conditions.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

As discussed in Section 2, our more sophisticated calculations of formation energies performed in the course of the present program confirmed that the results obtained in our original work [R5] are still valid. We can therefore rely on the calculated point-defect concentrations reported in Ref. [R36]. There it was found that, in the absence of any impurities, the nitrogen and gallium vacancies are the dominant defects, but their concentrations are still lower than 10^{15} cm^{-3} even at growth temperatures of 1100 °C. This indicates that we are still far from the point where point defects should become a dominant worry. Indeed, given that experimental techniques such as Secondary Ion Mass Spectroscopy (SIMS) are usually only capable of identifying impurities with a resolution of about 10^{16} cm^{-3}, and given the known tendency of GaN to readily incorporate background impurities such as O, C, and H, there is reason to believe that control of impurities should, for now, remain the primary focus of efforts to control the conductivity of GaN.

9. "Fingerprints" of defects and impurities to assist in experimental identification

Computations can provide various types of information to enable experimental identification of defects. The "signatures" we focused on in the course of the present program were frequencies of local vibrational modes, which can be measured with infrared or Raman spectroscopy, and the position of defect levels in the gap, to be compared with photoluminescence. Examples of both were mentioned throughout this report. For instance, in Section 1.a. we reported that we have calculated local vibrational modes for a large number of configurations that hydrogen can assume in GaN. In Ref. [P14] we address isolated interstitial hydrogen, Mg-H complexes, and Be-H complexes, in all their possible configurations. Most of these configurations involve N-H bonds, and based on these investigations of a large number of configurations we were able to derive a correlation between the vibrational frequency of the stretching mode and the bond length in the N-H bond. In Ref. [P15] we address the vibrational frequency of the H_2 molecule in wurtzite GaN, finding that it is red-shifted from the free molecule; for wurtzite GaN the frequency is 131 cm^{-1} lower than in free H_2.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

10. Additional topics

10.a. Role of hydrogen in growth of GaN

These results were described in detail in our Progress Report for the periods August 1, 2000 – July 31, 2001; they have reported in publications [P4] and [P11]. Brief summary:

A detailed understanding of surface reconstructions on nitride semiconductors is essential for improved control over growth, materials properties, and interface formation. Hydrogen plays an important role in these processes, and is abundantly present in commonly used growth techniques for nitrides. We have performed first-principles calculations allowing us to investigate the stability of surface reconstructions as a function of stoichiometry (Ga-rich *vs.* N-rich conditions), as well as of the abundance of hydrogen in the growth environment. We find that the reconstructions that are stable at high temperatures (relevant for MOCVD growth) are very different from those that occur at T=0. N-rich conditions favor N-H bonds (which are overwhelmingly favored at T=0), but moving towards Ga-rich conditions we find that 2×2 reconstructions involving only Ga-H bonds (3/4 monolayer H coverage), or even hydrogen-free surfaces (Ga adatom or Ga bilayer reconstructions) are more stable.

Changes in temperature and pressure strongly affect the relative stability of various reconstructions; such information is best presented in a phase diagram, but due to the large number of degrees of freedom a traditional phase diagram would be difficult to analyze or visualize. We have been able to show [P4] that the number of degrees of freedom can be reduced to *two* if *chemical potentials* are chosen as the variables. The resulting generalized surface phase diagram provides immediate insight in the stability of various surface structures as a function of growth conditions. The power of the approach has been evidenced by its successful use in interpreting the *in situ* observations of surface reconstructions based on grazing incidence x-ray scattering by Munkholm *et al.* [R37]. This favorable comparison with experiment illustrates that we have developed a powerful computational approach for obtaining phase diagrams for systems consisting of an impurity-covered surface in thermodynamic contact with a gaseous environment, as characteristic of MOCVD or HVPE (hydride vapor phase epitaxy). The approach gives imme-

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

diate access to the atomic structure and phase transitions of surfaces under realistic growth conditions. We are confident that the methodology can be generalized to surface alloying or oxidation.

10.b. Adatom kinetics on and below the surface: existence of a new diffusion channel

One of the degrees of freedom available to growers of GaN is the relative amounts of Ga and N delivered to the surface. It happens that when the growing surface is too rich in nitrogen, the growth surface becomes rough. Much smoother surfaces are obtained when the surface is grown under Ga-rich conditions. The explanation proposed in Ref. [P16] goes under the heading *Adlayer Enhanced Lateral Diffusion*, or AELD. A large body of work has led to the conclusion that in Ga-rich conditions the GaN(0001) surface is covered by two or more layers of Ga. A similar type of structure occurs when In is employed as a surfactant. In that case the surface is terminated by an In adlayer on top of a Ga layer. Under these conditions N atoms will incorporate between the In and Ga layers. Calculations [P16] have been performed to show that such N atoms can diffuse much more rapidly between these two layers than they could diffuse in the absence of the In adlayer. The activation energy for N diffusion is 1.3 eV in the absence of the adlayer and 0.5 eV in the presence of the adlayer. Rapid nitrogen diffusion results in smooth growth. A paper describing the calculations has been submitted for publication [P16].

Personnel supported

Chris G. Van de Walle, John E. Northrup, Sukit Limpijumnong.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

References

[R1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964); W. Kohn and L. J. Sham, *ibid.* **140**, A1133 (1965).

[R2] R. Stumpf and M. Scheffler, Comp. Phys. Commun. **79**, 447 (1994).

[R3] J. Neugebauer and C. G. Van de Walle, Proc. Mater. Res. Soc. Symp. **408** (1996).

[R4] M. Bockstedte, A. Kley, J. Neugebauer, and M. Scheffler, Comp. Phys. Commun. **107**, 187 (1997).

[R5] J. Neugebauer and C. G. Van de Walle, Phys. Rev. B **50**, 8067 (1994).

[R6] J. Neugebauer and C. G. Van de Walle, Proc. Mater. Res. Soc. Symp. **339**, 687 (1994).

[R7] S. G. Louie, S. Froyen, M. L. Cohen, Phys. Rev. B **26**, 1739 (1982).

[R8] J. Neugebauer and C. G. Van de Walle, Appl. Phys. Lett. **68**, 1829 (1996).

[R9] B. Clerjaud, D. Côte, A. Lebkiri, C. Naud, J. M. Baranowski, K. Pakula, D. Wasik, and T. Suski, Phys. Rev. B **61**, 8238 (2000).

[R10] J. Neugebauer and Chris G. Van de Walle, J. Appl. Phys. **85**, 3003 (1999).

[R11] S. Nakamura, M. Senoh, and T. Mukai, Jpn. J. Appl. Phys., Part 2, **30**, L1708 (1991).

[R12] J. Neugebauer and C. G. Van de Walle, Mater. Res. Soc. Symp. Proc. 395, **645** (1996).

[R13] C. G. Van de Walle, Phys. Rev. B **56**, R10020 (1997).

[R14] T. Yamamoto and H. Katayama-Yoshida, in *Proceedings of the 19th International Conference on Defects in Semiconductors*, Aveiro, Portugal, 1997, edited by G. Davies and M. H. Nazaré, Mat. Sci. Forum **258-263** (Trans Tech, Zürich, 1997), p. 1185.

[R15] O. Brandt, H. Yang, H. Kostial, and K. H. Ploog, Appl. Phys. Lett. **69**, 2707 (1996).

[R16] R. Y. Korotkov, J. M. Gregie, and B. W. Wessels, Appl. Phys. Lett. **78**, 222 (2001).

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

[R17] K. H. Chow, G. D. Watkins, A. Usui, and M. Mizuta, Phys. Rev. Lett. **85**, 2761 (2000).

[R18] Q. Zhou and M. O. Manasreh, M. Pophristic, S. Guo, and I. T. Ferguson, Appl. Phys. Lett. **79**, 2901 (2001).

[R19] Q. Zhou and M. O. Manasreh, Appl. Phys. Lett. **80**, 2072 (2002).

[R20] J. Oila et al., Phys. Rev. B **63**, 045205 (2001).

[R21] K. Saarinen, T. Suski, I. Grzegory, and D. C. Look, Phys. Rev. B **64**, 233201 (2001).

[R22] J. Neugebauer and Chris G. Van de Walle, Appl. Phys. Lett. **69**, 503 (1996).

[R23] M. A. Reshchikov, H. Morkoç, S. S. Park, and K. Y. Lee, Appl. Phys. Lett. **78**, 3041 (2001).

[R24] L.T. Romano, M. D. McCluskey, C. G. Van de Walle, J. E. Northrup, D. P. Bour, M. Kneissl, T. Suski, and J. Jun, Appl. Phys. Lett. **75**, 3950 (1999).

[R25] Y. Xin, S. J. Pennycook, D. D. Browning, P. D. Nellist, S. Sivanathan, F. Omnes, B. Beaumont, J. P. Faurie, and P. Gibart, Appl. Phys. Lett. **72**, 2680 (1998).

[R26] J. W. P. Hsu, M. J. Manfra, S. N. G. Chu, C. H. Chen, L. N. Pfeiffer, and R. J. Molnar, Appl. Phys. Lett. **78**, 3980 (2001).

[R27] L. T. Romano, C. G. Van de Walle, J. W. Ager III, W. Götz, and R. S. Kern, J. Appl. Phys. **87**, 7745 (2000).

[R28] S. Porowski, J. Cryst. Growth **189/190**, 153 (1998).

[R29] P. Perlin, T. Suski, A. Polian, J. C. Chervin, W. Knap, J. Camassel, I. Grzegory, S. Porowski and J. W. Erickson, Mater. Res. Soc. Symp. Proc. **449**, 519 (1997).

[R30] C. Stampfl, C. G. Van de Walle, D. Vogel, P. Krüger, and J. Pollmann, Phys. Rev. B **61**, R7846 (2000).

[R31] "Quantitative analysis of the polarization fields and absorption changes in InGaN/GaN quantum wells with electroabsorption spectroscopy", F. Renner, P. Kiesel, G. H. Döhler, M. Kneissl, C. Van de Walle, and N. M. Johnson, Appl. Phys. Lett. (in press).

[R32] C. G. Van de Walle and J. Neugebauer, Appl. Phys. Lett. **76**, 1009 (2000).

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

[R33] T. H. Myers, private communication.

[R34] J. Neugebauer and C. G. Van de Walle, Phys. Rev. Lett. **75**, 4452 (1995).

[R35] D. C. Look, H. Lu, W. J. Schaff, J. Jasinski, and Z. Liliental-Weber, Appl. Phys. Lett. **80**, 258 (2002).

[R36] J. Neugebauer and C. G. Van de Walle, in *Festkörperprobleme/Advances in Solid State Physics*, Vol. **35**, ed. by R. Helbig (Vieweg, Braunschweig/Wiesbaden, 1996), p. 25.

[R37] A. Munkholm *et al.*, Phys. Rev. Lett. **83**, 741 (1999).

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

Publications

Published:

Journal articles

[P1] "Entropy-driven stabilization of a novel configuration for acceptor-hydrogen complexes in GaN", S. Limpijumnong, C. G. Van de Walle, and J. E. Northrup, Phys. Rev. Lett. **87**, 205505 (2001).

[P2] "First-principles studies of beryllium doping of GaN", C. G. Van de Walle, S. Limpijumnong, and J. Neugebauer, Phys. Rev. B **63**, 245205 (2001).

[P3] "Influence of microstucture on the carrier concentration of Mg-doped GaN films", L. T. Romano, M. Kneissl, J. E. Northrup, C. G. Van de Walle, and D. W. Treat, Appl. Phys. Lett. **79,** 2734 (2001).

[P4] "First-principles surface phase diagram for hydrogen on GaN surfaces", C. G. Van de Walle and J. Neugebauer, Phys. Rev. Lett. **88**, 66103 (2002).

[P5] "Passivation and doping due to hydrogen in III-nitrides", S. Limpijumnong and Chris G. Van de Walle, phys. stat. sol. (b) **228**, 303 (2001).

[P6] "Theoretical investigation of native defects, impurities and complexes in AlN", C. Stampfl and C. G. Van de Walle, Phys. Rev. B **65**, 155212 (2002).

[P7] "Screw dislocations in GaN: The Ga-filled core model", J. E. Northrup, Appl. Phys. Lett. **78**, 2288 (2001).

Conference Proceedings

[P8] "Controlling the conductivity of wide-band-gap semiconductors", C. G. Van de Walle and J. Neugebauer, in *Proceedings of the 25th International Conference on the Physics of Semiconductors*, Osaka, 2000, edited by N. Miura and T. Ando (Springer, Berlin, 2001), p. 3.

[P9] "Vibrational spectroscopy of GaN:Mg under pressure", M. D. McCluskey, K. K. Zhuravlev, M. Kneissl, W. Wong, D. Treat, S. Limpijumnong, C. G. Van de Walle, and N. M. Johnson, in *GaN and Related Alloys*, edited by J. E. Northrup, J.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

Neugebauer, S. F. Chichibu, D. C. Look, and H. Riechert, Materials Research Society Symposium Proceedings, Vol. 693, I2.4 (2002).

[P10] "Novel configuration of Mg-H complexes in GaN", S. Limpijumnong, J. E. Northrup, and C. G. Van de Walle, in *GaN and Related Alloys*, edited by J. E. Northrup, J. Neugebauer, S. F. Chichibu, D. C. Look, and H. Riechert, Materials Research Society Symposium Proceedings, Vol. 693, I2.5 (2002).

Accepted for publication:

[P11] "Role of hydrogen in surface reconstructions and growth of GaN", Chris G. Van de Walle and J. Neugebauer, J. Vac. Sci. Technol. B (in press).

[P12] "Theory of intrinsic and H-passivated screw dislocations in GaN", J. E. Northrup, Phys. Rev. B (in press).

Submitted:

[P13] "Effects of impurities on the lattice parameters of GaN", C. G. Van de Walle, submitted to Phys. Rev. B.

[P14] "Identification of configurations of hydrogen in *p*-type GaN through first-principles calculations of vibrational frequencies", Sukit Limpijumnong, John E. Northrup, and Chris G. Van de Walle, submitted to Phys. Rev. B.

[P15] "Stability, diffusivity and vibrational frequencies of interstitial hydrogen in wurtzite GaN", Sukit Limpijumnong and Chris G. Van de Walle, submitted to Phys. Rev. B.

[P16] "Adatom kinetics on and below the surface: The existence of a new diffusion channel", J. Neugebauer, M. Scheffler, T. Zywietz, J. E. Northrup, H. Chen and R. M. Feenstra, submitted to Phys. Rev. Lett.

In preparation:

[P17] "Diffusivity of native defects in GaN", S. Limpijumnong and C. G. Van de Walle, to be submitted to Phys. Rev. B.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

[P18] *Applied Physics Review*: "First-principles calculations for defects and impurities: applications to III-nitrides", C. G. Van de Walle and J. Neugebauer, to be submitted to J. Appl. Phys.

[P19] "Interactions of silicon with screw dislocations in GaN", J. E. Northrup (unpublished).

Interactions/Transitions

- C. G. Van de Walle attended the *Ψk 2000 Conference on "Ab initio calculations of complex processes in materials"*, Schwäbisch Gmünd, Germany, August 22-26, 2000, and presented an **Invited Talk** on "Properties of GaN surfaces: the role of hydrogen".
- C. G. Van de Walle attended the *25th International Conference on the Physics of Semiconductors*, Osaka, Japan, September 17-22, 2000, and presented a **Plenary Talk** on "Controlling the conductivity of wide-band-gap semiconductors".
- C. G. Van de Walle and S. Limpijumnong participated in the "**Wood/Witt Review Meeting**", Boston, Massachusetts, November 26, 2000. C. G. Van de Walle presented a talk on "Unintentional impurities in nitride semiconductors".
- C. G. Van de Walle attended the *Fall Meeting of the Materials Research Society*, Boston, MA, November 27-December 1, 2000. C. G. Van de Walle presented an **Invited Talk** on "Role of hydrogen in surface reconstructions and growth of GaN".
- C. G. Van de Walle presented a Seminar in the Department of Physics, University of California, Santa Cruz, February 16, 2001, entitled: "Defects and impurities in wide-band-gap semiconductors".
- C. G. Van de Walle presented a Seminar at the California Institute of Technology, February 28, 2001, entitled: "Controlling the conductivity of wide-band-gap semiconductors".

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

- C. G. Van de Walle attended the *2001 March Meeting of the American Physical Society*, Seattle, WA, March 12-16, 2001, and presented a talk on "Theoretical study of monatomic hydrogen in III-nitrides".
- C. G. Van de Walle attended the *Workshop on Doping Issues in Wide-Band-Gap Semiconductors*, Exeter, England, March 21-23, 2001, and presented an **Invited Talk** on "Dopant engineering in wide-band gap semiconductors".
- C. G. Van de Walle presented a Seminar in the Department of Materials Science, University of California, Berkeley, April 9, 2001, entitled: "Controlling the conductivity of wide-band-gap semiconductors".
- S. Limpijumnong and J. E. Northrup attended the *Fourth International Conference on Nitride Semiconductors*, Denver, CO, July 16-20, 2001. S. Limpijumnong presented a talk on "Passivation and doping due to hydrogen in III-Nitrides"; J. E. Northrup presented a talk on "Dislocation core reconstructions in GaN".
- C. G. Van de Walle presented a Seminar at the University of Konstanz, Germany, September 4, 2001, entitled "Defects, impurities and complexes in wide-band-gap semiconductors".
- C. G. Van de Walle presented a Seminar at the Fraunhofer Institute, Freiburg, Germany, September 6, 2001, entitled "Dopant engineering in wide-band-gap semiconductors".
- C. G. Van de Walle presented a Seminar in the Dept. of Materials Science and Engineering, University of Michigan, Ann Arbor, Michigan, October 5, 2001, entitled "Controlling the conductivity of wide-band-gap semiconductors".
- C. G. Van de Walle presented a Seminar at Lumileds, San Jose, November 15, 2001, entitled "Role of hydrogen in growth of GaN".
- C. G. Van de Walle, John E. Northrup, and S. Limpijumnong participated in the "**Wood/Witt Review Meeting**", Boston, Massachusetts, November 25, 2001. C. G. Van de Walle presented a talk on "Point defects and impurities in GaN".
- C. G. Van de Walle, John E. Northrup, and S. Limpijumnong attended the *Fall Meeting of the Materials Research Society*, Boston, MA, November 26-30, 2001.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

C. G. Van de Walle presented a talk on "Novel configuration for acceptor-hydrogen complexes in GaN". S. Limpijumnong presented a talk on "Diffusivity of native defects in wurtzite GaN". John E. Northrup presented a talk on "Structure and stoichiometry of screw dislocations in GaN: theory".

- Chris G. Van de Walle attended the *29th International Conference on Physics in Semiconductors*, Santa Fe, New Mexico, January 6-10, 2002, and presented an **Invited Talk** on "Role of hydrogen in surface reconstructions and growth of GaN".
- John E. Northrup participated in the *ONR Workshop on Extended Defects in Wide Gap Semiconductors*, Belize, January 27-31, 2002, and present an **Invited Talk** on "Stoichiometry and electronic structure of screw dislocations in GaN.
- Chris G. Van de Walle and John E. Northrup attended the March Meeting of the American Physical Society, Indianapolis, Indiana, March 18-22, 2002. Chris G. Van de Walle presented the **Adler Award Lecture:** "The fascinating physics of hydrogen in semiconductors". John E. Northrup presented an **Invited Talk** on "Electrical activity of dislocations and extended defects in GaN".
- Chris G. Van de Walle attended the *19th General Conference of the Condensed Matter Division of the European Physical Society*, Brighton, United Kingdom, April 7-11, 2002, and presented an **Invited Talk** on "Defects and doping in wide-band-gap semiconductors".
- C. G. Van de Walle presented a Seminar at the University of Aarhus, Denmark, May 29, 2002, entitled "The fascinating physics of hydrogen in semiconductors".
- C. G. Van de Walle presented a Seminar at the University of Erlangen, Germany, June 10, 2002, entitled "Controlling the conductivity of wide-band-gap semiconductors".
- Chris G. Van de Walle attended the *11th International Conference on Metal-Organic Vapour Phase Epitaxy*, Berlin, Germany, June 3-7, 2002, and presented an **Invited Talk** on "Structure and energetics of nitride surfaces under MOCVD growth conditions".

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

New discoveries, inventions, patent disclosures

None

Honors/Awards

- Chris G. Van de Walle was promoted to the rank of **Principal Scientist** at the Xerox Palo Alto Research Center.
- Chris G. Van de Walle received the **David Adler Award** of the American Physical Society, 2002.
- Chris G. Van de Walle's collaboration with scientists at the Fritz-Haber-Institute and the Paul-Drude-Institut, Berlin, is supported by a **US Senior Scientist Award** from the Alexander von Humboldt Foundation.

Limited Rights
Contract No. F49620-00-C-0019
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

Screw dislocations in GaN: The Ga-filled core model

John E. Northrup[a)]
Xerox PARC, 3333 Coyote Hill Road, Palo Alto, California 94304

(Received 22 January 2001; accepted for publication 12 February 2001)

First-principles total energy calculations performed for [0001] screw dislocations in GaN with $|\mathbf{b}| = c$ indicate that a model with a helical Ga-filled core is more stable than the hollow core model in Ga-rich conditions. This model gives rise to electronic states dispersed throughout the band gap. Such a dislocation is therefore expected to be a very strong center for nonradiative recombination and a pathway for current leakage. © *2001 American Institute of Physics.*
[DOI: 10.1063/1.1361274]

There is considerable interest in determining the atomic structures and electronic properties of the various types of threading dislocations that are present in epitaxially grown GaN films. Therefore experimental[1,2] and theoretical studies[3–5] have been performed to ascertain these properties. These studies are motivated in part by a desire to understand the extent to which the various kinds of threading dislocations give rise to nonradiative recombination and affect charge carrier mobility. It is now well established that the performance of lasers improves as the density of dislocations is decreased.[6] For example, there is evidence that the photoluminescence intensity[7] and operating lifetime of lasers[8] increase as the densities of screw and mixed dislocations are decreased. In this letter we introduce a structural model for the pure screw dislocation that is energetically favorable in Ga-rich conditions and gives rise to electronic states throughout the band gap of GaN.

The two simplest types of threading dislocations are the pure screw, with Burgers vector $\mathbf{b}=[0001]$, and the pure edge dislocation, with $\mathbf{b}=1/3[11\bar{2}0]$. A projected atomic image of the edge dislocation has been obtained by Xin *et al.*[2] using transmission electron microscopy (TEM). This image is in accord with the atomic model predicted theoretically by Elsner *et al.*[3] In that model the core of the dislocation contains pairs of threefold coordinated Ga and N atoms. As on the $(10\bar{1}0)$ surface, which has a locally similar atomic structure, there exists a large energy gap between the occupied and empty dislocation induced states. Consequently, it has been suggested that pure edge dislocations are not intrinsic sources of nonradiative recombination events. However, the stress field produced by the dislocation may give rise to an accumulation of electrically active impurities and point defects near these dislocations.[3]

The end-on TEM image obtained by Xin *et al.*[2] for the screw dislocations seems to imply a filled core. An interpretation of this observation in terms of a stoichiometric full core dislocation is in conflict with the theoretical result of Elsner *et al.* that a small open core dislocation having a diameter of 7 Å is preferred energetically by a wide margin over the filled core model.[3] In this work we provide a resolution to this problem. Specifically we find that a screw dislocation having a core comprised entirely of Ga atoms is preferred energetically over the hollow core model in Ga-rich conditions. The structure is obtained by removal of all nitrogen atoms within 2 Å of the center of the dislocation core, i.e., six N atoms per unit cell along the c axis (~5.1 Å) are removed. For this structure electronic states are distributed throughout the band gap. Therefore such a model predicts that screw dislocations will be very strong centers for nonradiative recombination and electrical conduction in material grown under Ga-rich conditions. We will refer to the new structure as the Ga-filled core model.

To determine the relative stability of different types of core reconstructions, total energy calculations[9] were carried out using density functional theory and the first-principles pseudopotential method.[10] This approach has been applied previously to extended defects[11] and surfaces[12] of GaN. The screw dislocation with line direction [0001] and having a Burgers vector $\mathbf{b}=[0001]$ is modeled using a supercell approach. Each cell contains one dislocation and the repeat distance along the c axis is equal to the magnitude of the Burgers vector (c=5.1 Å). A schematic representation [projected onto the (0001) plane] of the atoms within the supercell is shown in Fig. 1(a). Each cell contains 54 Ga and 54 N atoms. In addition, the dangling bonds on the surface atoms at the six equivalent {10-10} edges of the cell are saturated by 36 H atoms. A screw dislocation is generated by applying a displacement in the c direction to each atom in the cell. The vector field displacement that generates the screw dislocation is given by $(0,0,\Delta z_i)$ where

$$\Delta z_i = c/2\pi \tan^{-1}(y_i/x_i),$$

and x_i and y_i are the x and y coordinates of the ith atom. Starting from these initial positions the atoms are allowed to relax and the total energy is minimized. The plane wave cutoff used in these calculations is 40 Ry and the Ga pseudopotential is treated using the nonlinear core approximation.[13] The **k**-point sampling is performed with two to four **k** points distributed uniformly along the dislocation line direction.

We may think of the system as an infinitely long prism-shaped crystal of GaN where the six {10-10} surfaces of the crystal are passivated by H atoms and one screw dislocation runs along the center of the prism. The energy to form a dislocation in the system is obtained by comparing directly the total energies of systems with and without the screw dislocation. The formation energy per cell will increase loga-

a)Electronic mail: northrup@parc.xerox.com

 © 2001 American Institute of Physics
Downloaded 21 Jun 2002 to 141.14.132.4. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/aplo/aplcr.jsp

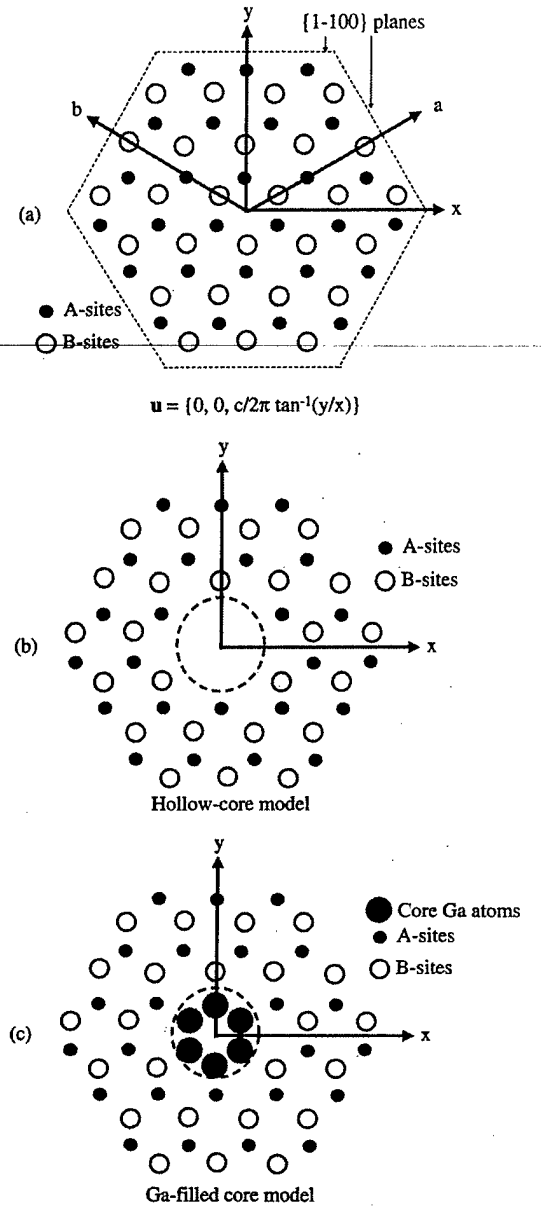


FIG. 1. (a) (0001) projection of the supercell employed to model a screw dislocation. The closed circles represent Ga and N atoms in the A sites of the wurtzite lattice. The open circles represent Ga and N atoms in B sites. Not shown are the H atoms which saturate the 18 Ga and 18 N dangling bonds in each cell on the six {10-10} surfaces. Each supercell contains one prismatic crystal of GaN. The supercell vectors **a** and **b** give rise to a periodic hexagonal arrangement of prisms in the x–y plane. The structure is periodic in the z direction with periodicity $c=5.1$ Å. The screw dislocation is generated by applying the displacement $(0,0,\Delta z_i)$. (b) A stoichiometric hollow core dislocation is created by removing the Ga and N atoms from within the central core of the filled core dislocation. (c) The Ga-filled core is obtained by removing the nitrogen atoms from the central core. The Ga atoms in the core form a pair of helices, each with pitch $2c$.

rithmically with the diameter of the system because of the long-range strain field produced by the screw dislocation. For the present case the diameter is ~15 Å and the calculated formation energy of a full core dislocation is 4.0 eV/Å. This result is similar to that obtained by Elsner *et al.* for systems of a similar size.[3]

Several types of dislocation core structures were considered. These include the stoichiometric full core model and the empty core models studied previously[3] and the Ga-filled core model obtained by removing all of the N atoms within

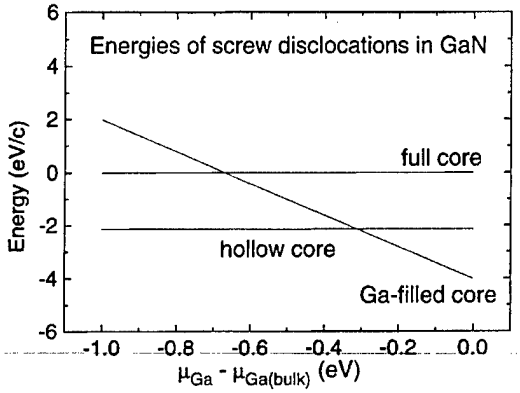


FIG. 2. Relative energies of possible screw dislocations as a function of the Ga chemical potential. The upper limit on μ_{Ga} is $\mu_{Ga(bulk)}$. The Ga-filled core mode is energetically favorable with respect to the hollow core model for $\mu_{Ga}-\mu_{Ga(bulk)}>-0.31$ eV.

an ~2 Å radius of the center of the dislocation. These models are indicated schematically in Fig. 1. The model is motivated in part by previous studies of surface reconstructions of GaN showing that Ga-terminated structures are stable in Ga-rich conditions.[12,14] Additional motivation for the Ga-filled core model stems from the fact that the nondirectional nature of metallic Ga–Ga bonding will render the energy much less sensitive to the enormous shear strains present at the core of the screw dislocation.

As for surfaces,[15] the equilibrium structure of a dislocation in a binary compound will depend, in general, on the chemical potentials of the two constituents, in this case on μ_{Ga} and μ_N. Since the sum of the two potentials is fixed by equilibrium with bulk GaN, we can express the relative formation energies of two different core reconstructions in terms of the Ga chemical potential alone. The difference in formation energies between dislocation structures containing different numbers of Ga and N atoms is

$$\Delta E=\Delta E(\text{Ga rich})-(\Delta n_{Ga}-\Delta n_N)(\mu_{Ga}-\mu_{Ga(bulk)}).$$

In this expression ΔE(Ga rich) is the calculated difference in energies corresponding to the Ga-rich limit, in which case $\mu_{Ga}=\mu_{Ga(bulk)}$. For equilibrium conditions $\mu_{Ga(bulk)}$ is the maximum possible chemical potential of Ga. The relative energies of three dislocation structures are plotted in Fig. 2 as a function of $(\mu_{Ga}-\mu_{Ga(bulk)})$. In agreement with previous work we find that the hollow core model is energetically favorable with respect to the full core model. These two structures have the same stoichiometry and so their relative energy is independent of the atomic chemical potentials. The energy difference is 2.14 eV/cell in the present calculations compared to 1.7 eV/cell obtained previously.[3] As shown in Fig. 2, the Ga-filled core model is found to be stable in Ga-rich conditions. It is preferred over the hollow core model for $\mu_{Ga}-\mu_{Ga(bulk)}>-0.31$ eV, and in the Ga-rich limit it is preferred by 1.87 eV/cell. On the basis of these calculations we propose this model for screw dislocations existing in GaN grown under Ga-rich conditions. This prediction is consistent with the end-on TEM images of screw dislocations obtained by Xin *et al.* that indicate some type of filled core structure.[2] According to our calculations the projected (0001) atomic positions of the Ga atoms in the Ga-filled core model remain within about 0.3 Å of those corre-

Downloaded 21 Jun 2002 to 141.14.132.4. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/aplo/aplcr.jsp

sponding to the stoichiometric filled core model. Thus, if the end-on TEM image is not sensitive to the reduced N concentration in the core, it will be difficult to distinguish between the two models. Clearly, further experimental studies are desirable to test the validity of the core structure proposed in the present work.

The Ga atoms in the core form a pair of helices, with each Ga atom in a helix bonded to two other Ga atoms and one N atom. The pitch of each helix is $2c$. The Ga–Ga bonds in the helix are 2.44 Å and the Ga–N bonds between the helix and the outlying N atoms are 1.90 Å in length. In projection onto the (0001) planes the Ga core atoms form a hexagon as shown in Fig. 1(c). The calculations predict that this hexagon is rotated 6° with respect to the rest of the lattice. Viewing along the [000-1] direction, the rotation is seen to be counterclockwise. In principle, such a rotation could be observable in a plan view TEM image.

Analysis of the Kohn–Sham eigenvalue spectrum for the Ga-filled core reveals a metallic electronic structure with states dispersed throughout the entire band gap of GaN. The Fermi level is located near midgap, about 1.4 eV above the valence band maximum. Such a dislocation would therefore be expected to be a very strong center for nonradiative recombination via the emission of phonons and Auger processes. In addition, in view of its metallic nature, it would be expected to give rise to a conductive pathway along the dislocation line. Recent experimental work of Hsu *et al.* indicates that screw dislocations provide a pathway for conduction in material grown by molecular beam epitaxy under Ga-rich conditions.[16]

In undoped material the line charge of the dislocation is zero. In doped material the GaN around the dislocation will be depleted and a small density of electrons/holes will accumulate along the core of the dislocation. The resulting band bending and intracore Coulomb interaction align the Fermi level of the dislocation with that in the bulk. For sufficiently high doping densities the charge transfer between nearby dopants and the dislocation could give rise to changes in the structure of the core. Decoration by impurities such as H (Ref. 5) or O (Ref. 17) would also affect the core structure and electrical activity.

In summary, we have found using first-principles total energy calculations that a Ga-filled core structure is much more stable than the hollow core structure under Ga-rich conditions. This screw dislocation gives rise to electronic states throughout the band gap of GaN, and hence would be a strong center for nonradiative recombination as well as for conduction along the dislocation line.

This work was supported in part by the AFOSR.

[1] S. J. Rosner, E. C. Carr, M. J. Ludowise, G. Girolami, and H. I. Erikson, Appl. Phys. Lett. **70**, 420 (1997).
[2] Y. Xin, S. J. Pennycook, N. D. Browning, P. D. Nellist, S. Sivanathan, F. Omnès, B. Beaumont, J. P. Faurie, and P. Gibart, Appl. Phys. Lett. **72**, 2680 (1998).
[3] J. Elsner, R. Jones, P. K. Sitch, V. D. Porezag, M. Elstner, Th. Frauenheim, M. I. Heggie, S. Öberg, and P. R. Briddon, Phys. Rev. Lett. **79**, 3672 (1997).
[4] A. F. Wright and U. Grossner, Appl. Phys. Lett. **73**, 2751 (1998).
[5] J. E. Northrup, R. Di Felice, and J. Neugebauer, Phys. Rev. B **56**, 4325 (1997).
[6] M. Hansen, P. Fini, L. Zhao, A. C. Abare, L. A. Coldren, J. S. Speck, and S. P. DenBaars, Appl. Phys. Lett. **76**, 529 (2000).
[7] T. Hino, S. Tomiya, T. Miyajima, K. Yanashima, S. Hashimoto, and M. Ikeda, Appl. Phys. Lett. **76**, 3421 (2000).
[8] T. Miyajima (private communication).
[9] R. Stumpf and M. Scheffler, Comput. Phys. Commun. **79**, 447 (1994).
[10] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[11] J. E. Northrup, J. Neugebauer, and L. T. Romano, Phys. Rev. Lett. **77**, 103 (1996).
[12] J. E. Northrup, J. Neugebauer, R. M. Feenstra, and A. R. Smith, Phys. Rev. B **61**, 9932 (2000).
[13] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).
[14] A. R. Smith, R. M. Feenstra, D. W. Greve, J. Neugebauer, and J. E. Northrup, Phys. Rev. Lett. **79**, 3934 (1997). On the GaN(000-1) surface, the surface is terminated by one monolayer of Ga, with each Ga atom in the adlayer bonded to a single N atom. This structure is stabilized by a strong interaction between neighboring surface Ga atoms.
[15] G. X. Qian, R. M. Martin, and D. J. Chadi, Phys. Rev. B **38**, 7649 (1988); J. E. Northrup, Phys. Rev. Lett. **62**, 2487 (1989).
[16] J. W. P. Hsu, M. J. Manfra, D. V. Lang, S. Richter, S. N. G. Chu, A. M. Sergent, R. N. Kleiman, L. N. Pfeiffer, and R. J. Molnar, Appl. Phys. Lett. **78**, 1685 (2001).
[17] J. Elsner, R. Jones, M. Haugk, R. Gutierrez, Th. Frauenheim, M. I. Heggie, S. Öberg, and P. R. Briddon, Appl. Phys. Lett. **73**, 3530 (1998); Z. Liliental-Weber, Y. Chen, S. Ruvimov, and J. Washburn, Phys. Rev. Lett. **79**, 2835 (1997).

Downloaded 21 Jun 2002 to 141.14.132.4. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/aplo/aplcr.jsp

Influence of microstructure on the carrier concentration of Mg-doped GaN films

L. T. Romano,[a] M. Kneissl, J. E. Northrup, C. G. Van de Walle and D. W. Treat
Xerox Palo Alto Research Center, Electronic Materials Laboratory, 3333 Coyote Hill Road, Palo Alto, California 94304

(Received 23 July 2001; accepted for publication 27 August 2001)

Room-temperature Hall effect measurements of (0001) Mg-doped GaN films grown on sapphire substrates by metalorganic chemical vapor deposition show a reduction in hole concentration for Mg concentrations greater than $10^{20}\,cm^{-3}$. A combination of secondary ion mass spectrometry and transmission electron microscopy indicates a steadily increasing Mg incorporation during growth and the formation of inversion domains at these high concentrations. We discuss mechanisms that could give rise to a reduction of the hole concentration at high Mg doping levels. © *2001 American Institute of Physics.* [DOI: 10.1063/1.1413222]

Magnesium is the most common acceptor in *p*-type GaN alloys and is widely used in GaN-based light emitting and laser diodes. However, Mg has a relatively large ionization energy ($E_A \sim 180$ meV), which limits the maximum hole carrier concentration that can be achieved in Mg-doped GaN films. For example, at 300 K, less than 1% of the Mg acceptors are ionized with the highest reported hole concentration of $\sim 10^{18}\,cm^{-3}$ for a Mg concentration [Mg] of $\sim 10^{20}\,cm^{-3}$.[1] Unfortunately, it has been found that at higher [Mg], the hole carrier concentration decreases and surface roughness develops for films grown by metalorganic chemical vapor deposition (MOCVD).[2,3]

Recent work by Ramachandran *et al.* shows that Mg on the surface of Ga-polar (0001) GaN induces the formation of inversion domains (ID) on the (0001) basal plane.[4] Magnesium segregation and the formation of an inversion domain boundary (IDB) was found when the Mg concentration at the surface exceeded 1 ML for films grown by molecular beam epitaxy.[5,6] In this case the IDB was faceted and continuous along the interface between the layers of Ga and N polarity and found to initiate only on Ga-polar surfaces.[5] A model was proposed in which Mg resides on the faceted boundary in threefold coordinated Ga sites[5] and therefore would not act as a *p*-type dopant. In this letter, a structural study is made of films grown by MOCVD with various Mg concentrations and compared to their carrier concentration.

A series of ~1.5-μm-thick Mg-doped GaN films with different acceptor concentrations was grown on 4 μm of unintentionally doped *n*-type GaN on sapphire substrates by MOCVD. The Mg doping was controlled by varying the Cp_2Mg flow during growth. The [Mg] was measured by secondary ion mass spectrometry (SIMS) and found to track the relative Cp_2Mg flow rates. After growth, the Mg-doped films were annealed at 850 °C for 5 min in N_2 atmosphere in order to activate the Mg acceptors. Room-temperature Hall effect measurements with a Van der Pauw geometry were performed to determine hole carrier concentrations.

The full width at half maximum (FWHM) of the (0006) diffraction peak was measured by high-resolution x-ray diffraction (XRD) and found to increase with [Mg], ranging from 5.3 to 6.5 minutes for $[Mg] = 1\times10^{20}$ and $1.8\times10^{20}\,cm^{-3}$, respectively. The FWHM was found to be slightly greater for unintentionally doped films having typical values of ~4 min. Transmission electron microscopy (TEM) was performed on films that were mechanically polished and ion milled to electron transparency. The polarity was determined by convergent beam electron diffraction (CBED) and the microstructure was analyzed by conventional diffraction contrast[7] and multiple (0002) dark field imaging.[8] The unintentionally doped *n*-type GaN layers were found to have Ga polarity. Cross-section scanning electron microscopy (SEM) was used to determine the thickness of the GaN:Mg layer from contrast differences due to doping.

The hole concentration for a series of *p*-doped GaN films is shown in Fig. 1 as a function of [Mg]. It can be observed that the hole concentration initially increases with Mg until the [Mg] is $\sim 10^{20}\,cm^{-3}$ (Sample B). The room-temperature hole concentration for sample B was $1.2\times10^{18}\,cm^{-3}$ with a mobility of 0.7 $cm^2/V\,s$. Further increase in the [Mg] results in a significant decrease of the hole concentration with a measured carrier concentration of $1.6\times10^{17}\,cm^{-3}$ and a similar mobility of 0.7 $cm^2/V\,s$ (sample A).

The [Mg] as a function of film thickness for samples A

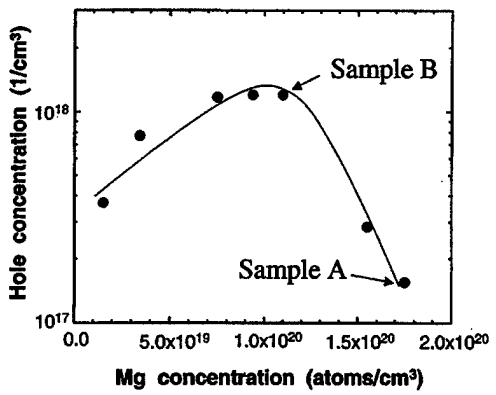


FIG. 1. Room-temperature Hall measurements of the hole concentration as a function of the Mg concentration (determined by SIMS) for a series of Mg-doped GaN films.

[a]Present address: Nova Crystals, San Jose, CA 95131; electronic mail: linda_romano@novacrystals.com

 © 2001 American Institute of Physics
Downloaded 23 Nov 2001 to 13.1.101.37. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/aplo/aplcr.jsp

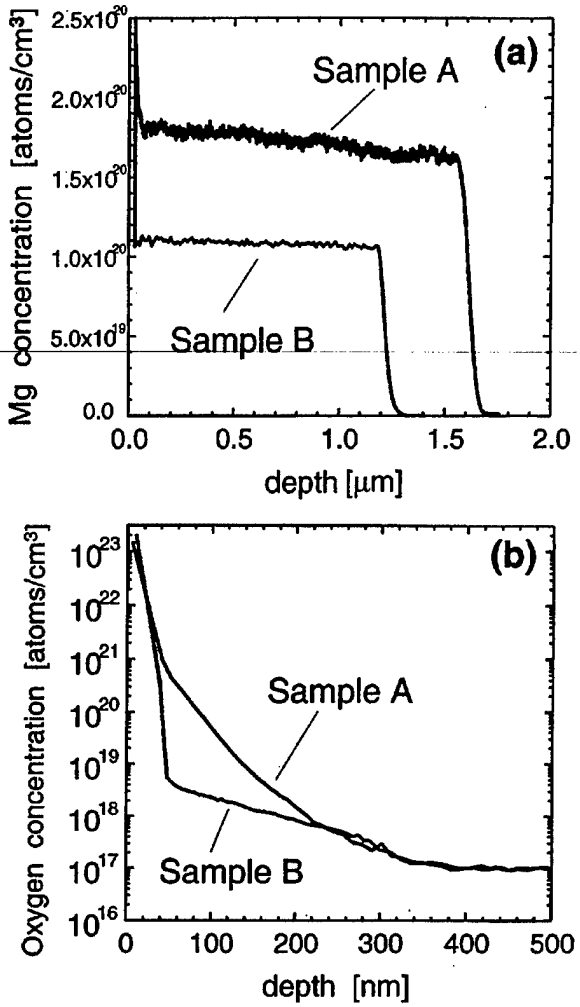


FIG. 2. SIMS depth profiles of the (a) Mg concentration and (b) O concentration for the high-doped (sample A) and medium-doped (sample B) GaN:Mg samples.

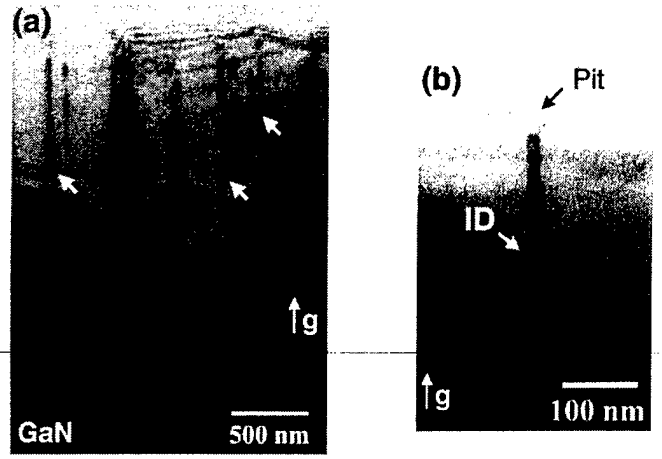


FIG. 3. TEM images of sample A taken with $g=0002$ near the [11$\bar{2}$0] zone axis. (a) Low magnification image showing the interface (dotted line) between the doped and the undoped layer. Inversion domains (ID) are indicated by arrows. Modulated strain contrast fringes are observed only in the doped layer. (b) High magnification image of an isolated ID associated with a pit at the film surface.

and B is shown in Fig. 2(a). For sample B, the [Mg] is $1.1\times10^{20}\,\mathrm{cm}^{-3}$ and remains constant through the doped portion of the sample. However, for sample A, the [Mg] increases from $1.6\times10^{20}\,\mathrm{cm}^{-3}$ at the doped/undoped interface to $1.8\times10^{20}\,\mathrm{cm}^{-3}$ at the surface. The measured O concentration [O] in these samples is shown in Fig. 2(b). Ignoring the initial surface layer of oxygen (depth$<$50 nm), the [O] background level for both samples was $\sim10^{17}\,\mathrm{cm}^{-3}$ and occurred at a depth of ~300 nm from the surface. The [O] in sample B abruptly decreased beyond the surface layer to $4\times10^{18}\,\mathrm{cm}^{-3}$ and approached the background level at ~1 decade/300 nm. However, for sample A, the [O] at the surface was much higher at $8\times10^{20}\,\mathrm{cm}^{-3}$ and the drop-off steeper at ~1 decade/100 nm. The high concentration of O near the surface of sample A is suggestive of Mg behaving as an O getter.

The microstructure of the films was observed by TEM. For films with $[\mathrm{Mg}]<1.1\times10^{20}\,\mathrm{cm}^{-3}$, the defect structure and polarity of the doped layer and undoped layers were similar. However, inversion domains and strain contrast were observed in doped films with a higher [Mg] as shown in Fig. 3 for sample B. The interface between the doped and nominally undoped layers is indicated by the dotted line in Fig. 3(a). The doped layer is ~1.5 μm thick which is consistent with growth-rate calibrations, cross-section SEM images, and the SIMS profile [Fig. 2(a)]. Modulated strain contrast is observed in the doped layer when imaged with the (0002) diffraction vector. The strain appears along the basal planes by alternating light/dark contrast bands that are separated by ~50–100 nm. The fringes are observed only in the doped layer and terminate at the doped/undoped layer interface. Similar strain contrast has been observed previously for highly doped Si:B layers and attributed to local bond distortions of the Si–B bond and B clustering.[9,10] The strain in our films, which is also evidenced by the increased FWHM in the XRD, may be due to the formation of Mg-related point defects, as discussed below.

Threading dislocations from the nominally undoped layer are observed to extend unperturbed through the interface into the doped layer. No additional threading dislocations, Mg precipitates or pyramidal defects[11] were observed in the Mg-doped layer. However, columnar defects along the {10$\bar{1}$0} planes were observed that initiated ~0.5 μm above the doped/undoped interface. These defects were found by (0002) dark field imaging and CBED to be IDs with N polarity at a density of $\sim10^{8}\,\mathrm{cm}^{-2}$. High-resolution TEM imaging (not shown) indicates that the domains nucleate on the basal plane which is consistent with previous studies of GaN:Mg films grown on Ga-polarity surfaces.[4,5] The high magnification image in Fig. 3(b) show that pits form where the IDs meet the surface. This is probably due to the different growth rates of the two polarities as reported previously.[12,13]

The reduction of the hole concentration that occurs when the total Mg concentration [Mg] exceeds 10^{20} suggests that some of the Mg incorporates in an electrically inactive form and/or gives rise to compensating donors at high Mg concentrations. First we examine whether the observed density and size of the IDs is sufficient to cause electrical deactivation of the material. From the TEM images the domains are, on average, separated by ~1 μm, and the typical diameter of the domain is ~30 nm. The ratio of atoms on the boundary to the total atom density is therefore $\sim5\times10^{-5}$, corresponding to a total density of boundary atoms of $\sim10^{18}/\mathrm{cm}^{3}$. If there is a donor defect associated with a significant fraction

Downloaded 23 Nov 2001 to 13.1.101.37. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/aplo/aplcr.jsp

of these boundary sites, then it is possible that the reduction in hole concentration (which is on the order of $10^{18}/\text{cm}^3$) can be attributed in part to donors associated with the ID boundaries. Studies of the interaction of Mg atoms with $\{10\underline{1}0\}$ ID boundaries are required in order to assess further this possibility. In the absence of decoration by impurities, the $\{10\underline{1}0\}$ ID boundaries are not expected to give rise to any states in the gap that would compensate *p*-type doping.[14]

The high oxygen concentration near the surface may also compensate some of the Mg. Simulations were performed by taking into account partial compensation of the Mg acceptors.[15] The difference in O background concentration between sample A and sample B was shown to reduce the hole concentration in sample A by ~20% (averaged over the 300 nm from the surface of the GaN:Mg film). However, this effect would be negligible since the thickness of the O-rich layer is only ~20% of the doped layer.

Finally, we consider the formation of compensating point defects. Most of the Mg is incorporated during MOCVD growth in the form of a neutral Mg–H complex, and these Mg atoms act as acceptors only after activation.[16] In order to account for a reduction in the hole concentration, we posit another type of Mg-related complex that would act as a *donor* after activation. There are many possibilities and we will discuss one example to illustrate the concept. Consider a complex consisting of a hydrogenated nitrogen vacancy[17] in which two of the surrounding Ga atoms have been replaced by Mg. We will refer to this complex as the $\text{Mg}_2-V_\text{N}-\text{H}$ complex. Such a complex would be a *neutral* defect during growth but would become a *donor* once the hydrogen was removed by the activation process. If we assume that equilibrium conditions prevail during the MOCVD growth process, the concentrations of these two complexes can be expressed in terms of their chemical-potential-dependent formation energies. Because these are neutral complexes the electron Fermi level does not enter, and we may write

$$\Omega(\text{Mg}-\text{H}) = \Omega_1 = E_1 - \mu_\text{Mg} + \mu_\text{Ga} - \mu_\text{H},$$

$$\Omega(\text{Mg}_2-V_\text{N}-\text{H}) = \Omega_2 = E_2' - 2\mu_\text{Mg} + 2\mu_\text{Ga} + \mu_\text{N} - \mu_\text{H} = E_2 - 2\mu_\text{Mg} + \mu_\text{Ga} - \mu_\text{H}.$$

In the expression for $\Omega(\text{Mg}_2-V_\text{N}-\text{H})$ we absorb the sum $\mu_\text{Ga}+\mu_\text{N}$ (which is constant) into the definition of E_2. We assume that the concentrations of the two complexes are each determined by these formation energies. After activation, the effect of these complexes on the difference between the number of acceptors and donors is given by the difference in the two concentrations

$$\Delta C \propto N \exp[-\Omega_1/kT]\{1-\exp[(E_1-E_2+\mu_\text{Mg})/kT]\}.$$

The important point is that the magnitude and perhaps even the sign of ΔC is governed by the Mg chemical potential. When μ_Mg is low the Mg–H complex dominates, but when μ_Mg increases the concentration of the $\text{Mg}_2-V_\text{N}-\text{H}$ complex may become comparable to or larger than that of the Mg–H complex. This simple model illustrates that if defect complexes stabilized under Mg-rich conditions act as donors following activation, then it is possible to account for the existence of a maximum in the hole concentration as a function of [Mg].

In conclusion, inversion domains and a reduction in the hole concentration was found in GaN films when the bulk concentration of Mg was in excess of 10^{20} cm^{-3}. The ID formation was accompanied by an increase in strain, as observed by TEM and XRD, and an increase in the Mg incorporation rate, as observed by SIMS. The reduction in hole concentration may be attributed to the incorporation of Mg in electrically inactive form and/or the formation of compensating Mg-induced defects at high Mg concentrations.

The authors would like to acknowledge Werner Goetz at Lumileds, San Jose, CA, for helpful discussions and Jon Erickson for SIMS measurements performed at Accurel Systems, Sunnyvale, CA. This work was supported in part by the AFOSR under Contract No. F4920-00-C-0019.

[1] S. Nakamura, M. Senoh, and T. Mukai, Jpn. J. Appl. Phys., Part 2 **30**, L1708 (1991).
[2] D. Bour, M. Kneissl, W. Goetz, L. T. Romano, C. G. Van de Walle, J. E. Northrup, R. M. Donaldson, J. Walker, and N. M. Johnson, Mater. Res. Soc. Symp. Proc. **449**, 509 (1997).
[3] I. P. Smorchkova, E. Haus, B. Heying, P. Kozodoy, P. Fini, J. P. Ibbetson, S. Keller, S. P. DenBaars, J. S. Speck, and U. K. Mishra, Appl. Phys. Lett. **76**, 718 (2000).
[4] V. Ramachandran, R. M. Feenstra, W. L. Sarney, L. Salamanca-Riba, J. E. Northrup, L. T. Romano, and D. W. Greve, Appl. Phys. Lett. **75**, 808 (1999).
[5] L. T. Romano, J. E. Northrup, A. J. Ptak, and T. H. Myers, Appl. Phys. Lett. **77**, 2479 (2000).
[6] A. J. Ptak, T. H. Myers, L. T. Romano, C. G. Van de Walle, and J. E. Northrup, Appl. Phys. Lett. **78**, 285 (2001).
[7] P. B. Hirsch, A. Howie, R. Nicholson, D. Pashley, and M. Whelan, *Electron Microscopy of Thin Crystals* (Krieger, Malabar, FL 1977).
[8] L. T. Romano, J. E. Northrup, and M. A. O'Keefe, Appl. Phys. Lett. **69**, 2394 (1996).
[9] A. Valilonis, G. Glass, P. Desjardins, D. G. Cahill, and J. E. Greene, Phys. Rev. Lett. **82**, 4464 (1999); G. Glass (private communication).
[10] D. D. Perovic, G. C. Weatherly, R. F. Egerton, D. C. Houghton, and T. E. Jackman, Philos. Mag. A **63**, 757 (1991).
[11] P. Vennegues, M. Benaissa, B. Beaumont, E. Feltin, P. De Mierry, S. Dalmosso, M. Leroux, and P. Gibart, Appl. Phys. Lett. **77**, 880 (2000).
[12] Z. Liliental-Weber, H. Sohn, N. Newman, and J. Washburn, J. Vac. Sci. Technol. B **13**, 1578 (1995).
[13] L. T. Romano and T. H. Myers, Appl. Phys. Lett. **71**, 3486 (1997).
[14] J. E. Northrup, J. Neugebauer, and L. T. Romano, Phys. Rev. Lett. **77**, 103 (1996).
[15] W. Götz, R. S. Kern, C. H. Chen, H. Liu, D. A. Steigerwald, and R. M. Fletcher, Mater. Sci. Eng., B **59**, 211 (1999).
[16] J. Neugebauer and C. G. Van de Walle, Phys. Rev. Lett. **75**, 4452 (1995), and references therein.
[17] C. G. Van de Walle, Phys. Rev. B **56**, R10020 (1997).

Downloaded 23 Nov 2001 to 13.1.101.37. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/aplo/aplcr.jsp

Controlling the conductivity of wide-band-gap semiconductors

Chris G. Van de Walle[1], J. Neugebauer[2]

[1] Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, USA e-mail: vandewalle@parc.xerox.com
[2] Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14 195 Berlin-Dahlem, Germany

Abstract Wide-band-gap semiconductors often exhibit limitations in the ability to control n-type or p-type doping. A theoretical framework is presented for studying doping of semiconductors, with key parameters derived from first-principles calculations. The formalism is illustrated with examples for GaN and ZnO, where increased understanding can lead to improved control of doping through *defect and impurity engineering.*

1 Introduction

The ability to control n-type and p-type conductivity is essential for design and fabrication of electronic and optoelectronic devices. Such conductivity control has traditionally been very difficult in wide-band-gap semiconductors, and native point defects have often been invoked to explain these problems. We will describe the formalism that we have developed to address the energetics of defect and impurity incorporation, and the effects of these species on the electronic properties. While building on notions developed in the 1950s, the current formulation in terms of formation energies and chemical potentials allows more direct insight into the mechanisms that govern incorporation of intrinsic and extrinsic defects.

Progress in computational physics now allows explicit evaluation of formation energies and the resulting defect and impurity concentrations entirely from first principles, i.e., without any input from experiment. Such state-of-the-art calculations are based on density functional theory and pseudopotentials, and carried out in a supercell geometry [1]. These fundamental concepts will be illustrated with examples from our recent work on III-V nitride and II-VI oxide semiconductors. We will review some key results for GaN, showing that the "conventional wisdom" regarding the role of native defects in doping and compensation is not always correct. For many years, nitrogen vacancies were thought to be the major source of the n-type conductivity that is commonly observed in as-grown GaN; our computational results showed, however, that formation of nitrogen vacancies is energetically unfavorable in n-type GaN [2]. We will also discuss the challenge of p-type doping in III-nitrides, and various strategies for impurity engineering that are aimed at overcoming the doping limitations [3].

We will also present some recent results for ZnO, a semiconductor with a great variety of applications. Recently, ZnO has come to the forefront as a promising candidate for new optoelectronic and electronic devices, but this potential can only be fulfilled if the conductivity of the material can be brought under control. Our calculations elucidate the role of native defects – showing, for instance, that the oxygen vacancy is a deep, rather than a shallow donor. As in GaN, native point defects are unlikely to be the source of the widely observed unintentional n-type conductivity. As an alternative source of n-type doping, we will discuss the incorporation of hydrogen impurities [4]. Hydrogen acts as a donor in ZnO – in contrast to its amphoteric behavior in other semiconductors. These results provide important guidelines for optimizing dopant impurity incorporation and activation.

The theoretical approach described here is quite general in nature, and can be applied to any semiconductor (as well as other materials) in which impurities and defects play a role. The current applications to wide-band-gap semiconductors constitute a nice illustration of the power of the approach.

2 Methodology

2.1 First-principles calculations

First-principles computational theory has significantly contributed to the fundamental understanding of defects and impurities in semiconductors. Such calculations yield the following types of information:

- microscopic structure, including atomic relaxations of the host atoms, derived by minimizing the total energy with respect to the atomic coordinates
- charge densities, which provide insight into the chemical bonding
- wave functions, which can be used for calculating hyperfine parameters, allowing direct comparison with experimental results obtained, e.g., with electron paramagnetic resonance
- band structures, which contain information about the optical and electronic properties
- frequencies of local vibrational modes (derived by evaluating the energy change or forces when atoms are displaced from their equilibrium positions)
- total energies, which allow evaluation of the relative stability of various configurations.

In addition, we have developed the methodology for using the total energies to provide a direct measure of the abundance of a defect or impurity in the crystal, as described in Sect. 2.2.

The first-principles calculations used in our studies are founded on density-functional theory, using a supercell geometry and *ab initio* pseudopotentials [5]. It is

widely known that the density-functional approach underestimates the band gap; this may affect the results for point defects that have occupied electronic states in the gap. Significant improvements in the band gaps of wide-band-gap semiconductors have been obtained by introducing self-interaction corrections. In particular, we have used an approach based on self-interaction and relaxation-corrected pseudopotentials, as described in Refs. [6] and [7]. Further details and references for the computational approach can be found in Refs. [2] and [4].

2.2 Formalism for calculating defect and impurity concentrations

The equilibrium concentration of an impurity or point defect is given by

$$c = N_{\text{sites}} \exp^{-E^f/k_B T} \quad (1)$$

where E^f is the *formation energy*, N_{sites} is the number of sites (per unit volume) on which the defect or impurity can be incorporated, k_B is the Boltzmann constant, and T the temperature. Equation (1) shows that defects with a *high* formation energy will occur in *low* concentrations.

The formation energy is not a constant but depends on the growth conditions. Let us illustrate this with the example of a nitrogen vacancy in GaN. The formation energy of this defect is determined by the relative abundance of Ga and N atoms, as expressed by the chemical potentials μ_{Ga} and μ_{N}. If the nitrogen vacancy (V_{N}) is charged (as is expected when it is electrically active), the formation energy depends further on the Fermi level (E_F), which acts as a reservoir for electrons. Forming a nitrogen vacancy requires the removal of one N atom, which is placed in the thermodynamic reservoir for N atoms, with energy μ_{N}; the formation energy is therefore:

$$E^f(V_{\text{N}}^+) = E_{\text{tot}}(V_{\text{N}}^+) - E_{\text{tot}}(\text{GaN bulk}) + \mu_{\text{N}} + E_F. \quad (2)$$

First-principles calculations allow explicit derivation of $E_{\text{tot}}(V_{\text{N}}^+)$, the total energy of a system containing a nitrogen vacancy. Similar expressions apply to other impurities and to the various native point defects.

Equation 1 relies on an assumption of thermodynamic equilibrium, which can be expected to be satisified at the high temperatures at which bulk growth or metal-organic chemical vapor deposition (MOCVD) of wide-band-gap semiconductors is carried out. At lower temperatures, such as those used in molecular-beam epitaxy (MBE), deviations from equilibrium may occur.

2.3 Chemical potentials

We emphasized above that the chemical potentials are free parameters, reflecting the relative abundance of various sources in the growth system. However, the chemical potentials are subject to specific bounds. For instance, in the case of GaN the upper limit on the Ga chemical potential is given by $\mu_{\text{Ga}}=\mu_{\text{Ga[bulk]}}$; indeed, pushing the Ga chemical potential to higher values would result in precipitation of bulk Ga rather than growth of GaN. This upper limit on μ_{Ga} places a lower limit on μ_{N}, because in equilibrium: $\mu_{\text{Ga}} + \mu_{\text{N}} = E_{tot}[\text{GaN}]$ where $E_{tot}[\text{GaN}]$ is the total energy of a two-atom unit of bulk GaN, calculated for the structurally optimized wurtzite structure. For the N-rich case, the upper limit on μ_{N} is given by $\mu_{\text{N}}=\mu_{\text{N[N}_2\text{]}}$, i.e., the energy of N in an N_2 molecule at $T = 0$; this yields a lower limit on μ_{Ga}.

For ease of presentation, we will choose specific values for the atomic chemical potentials when plotting our results; however, a general case can always be addressed by referring back to Eq. (2).

The Fermi level E_F is not an independent parameter, but is always determined by the condition of charge neutrality. In principle, equations such as (2) can be formulated for every native defect and impurity in the material; the complete problem (including free-carrier concentrations in valence and conduction bands) can then be solved self-consistently, imposing charge neutrality. However, it is instructive to plot formation energies as a function of E_F in order to examine the behavior of defects and impurities when the doping level changes.

The formalism described here is in principle equivalent to writing down mass-action relations for all defect reactions, as discussed, e.g., in Ref. [8]. However, the use of formation energies and chemical potentials (rather than partial pressures) renders the process amenable to general solution. In addition, the recent advances in computational physics now allow derivation of the key parameters from first principles. Indeed, the state-of-the-art calculations used to derive E_{tot} in Eq. (2) do not require any adjustable parameters or any input from experiment.

3 Results for GaN

3.1 Native defects and n-type doping

Figure 1 shows calculated formation energies for all native point defects in GaN, as a function of Fermi level, for Ga-rich conditions. The zero of E_F is located at the top of the valence band. For each defect, only the line segment is shown that corresponds to the charge state that gives rise to the lowest energy at a particular value of E_F. The change in slope of the lines therefore represents a change in the charge state of the defect [see Eq. (2)], and the Fermi-level position at which this change occurs corresponds to a transition level that can be experimentally measured.

Two major conclusions can immediately be drawn:

1. In thermodynamic equilibrium self-interstitials and antisites are too high in energy to form in significant concentrations. Only vacancies have sufficiently low energies to play a role in the electrical properties of the material.
2. Nitrogen vacancies have a high formation energy in n-type GaN; the corresponding concentration is too low to affect the electrical conductivity. Nitrogen vacancies can thus not be responsible for the commonly observed unintentional n-type conductivity in GaN.

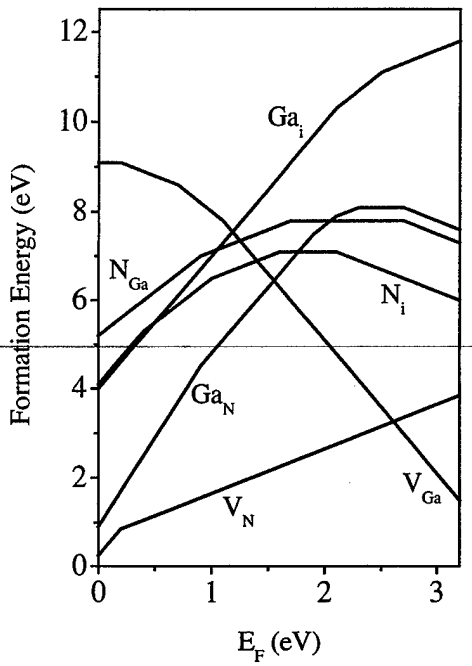


Fig. 1 Formation energies as a function of Fermi level for native point efects in GaN under Ga-rich conditions. $E_F = 0$ corresponds to the top of the valence band.

We have, instead, attributed this conductivity to the unintentional incorporation of donor impurities, such as silicon or oxygen [9]. These insights allow better control of the n-type doping of GaN, providing a good starting point for the achievement of p-type doping.

3.2 Hydrogen and p-type doping

Magnesium is currently the main p-type dopant used for nitrides. Unfortunately, its large ionization energy (210 meV, see Ref. [10]) poses severe limitations. About 10^{20} cm^{-3} Mg acceptors can be incorporated in GaN, but at room temperature, the hole concentration is only on the order of 10^{18} cm^{-3}. Increasing the Mg concentration does not improve the conductivity; in fact, it leads to lower hole concentrations [11]. We have performed extensive investigations in order to understand this behavior.

Hydrogen is known to play an important role in p-type doping of GaN [12–14]; indeed, many of the common growth techniques for GaN introduce large quantities of hydrogen into the growth environment. We have therefore devoted particular attention to the properties of interstitial hydrogen. The calculated formation energies for H in GaN are shown in Fig. 2.

Hydrogen prefers the positive charge state in p-type material, acting as a donor; it prefers the negative charge state in n-type material, acting as an acceptor. Hydrogen is thus an amphoteric impurity and always counteracts the prevailing conductivity of the semiconductor. This is similar to the behavior of hydrogen in other semiconductors such as silicon [15] and GaAs [16].

Let us now focus on the behavior in p-type GaN. Figure 3 displays formation energies of defects and im-

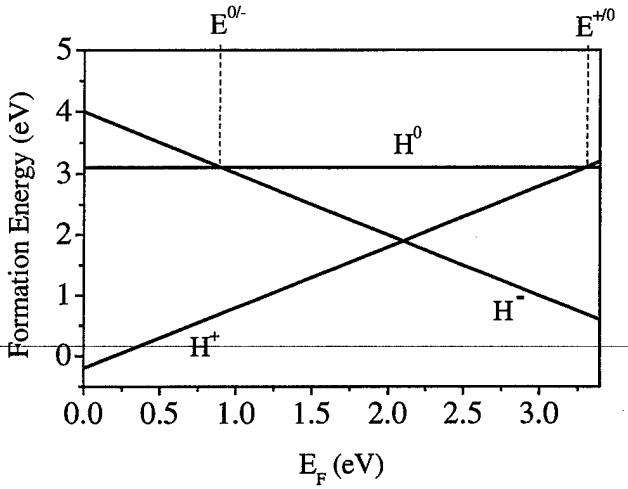


Fig. 2 Formation energies of various charge states of hydrogen in GaN as a function of Fermi level. $E_F = 0$ corresponds to the top of the valence band. The chemical potential of hydrogen was chosen to correspond to H_2 molecules at T=0.

purities relevant for p-type doping. The formation energy of Mg_{Ga} (the acceptor species) was determined by assuming equilibrium with Mg_3N_2; this corresponds to the solubility limit, i.e., the maximum impurity concentration that can be obtained in thermodynamic equilibrium. Trying to push the Mg concentration in GaN beyond this limit may result in precipitation of Mg_3N_2, which may be responsible for a decrease in crystal quality and the observed reduction in hole concentrations [11]. Incorporation of Mg on interstitial sites (Mg_i) or antisites (Mg_N), which had been suggested as an explanation for the limited effectiveness of Mg doping, is unlikely to be a problem, given the high formation energies of these species.

We note that compensation by nitrogen vacancies may be a problem in p-type GaN. Nitrogen vacancies have high formation energies in n-type GaN, but their donor character leads to a significant reduction of the formation energy in p-type material. In the presence of hydrogen, however, compensation by nitrogen vacancies will be suppressed. Indeed, the formation energy of hydrogen is lower than that of V_N, and hence hydrogen is more likely to incorporate as the compensating donor. Another beneficial effect due to hydrogen is an increase in the solubility of the acceptor. Imposing the condition of charge neutrality leads to the condition that the concentrations (and hence the formation energies) of acceptor and donor species must be equal at the growth temperature (at least when E_F is far enough from the band edges to allow neglect of free carriers). We see that the presence of hydrogen leads to a Fermi-level position higher in the band gap than in the absence of hydrogen; this leads to a decrease in the formation energy of Mg_{Ga}, and hence a higher solubility.

After the growth, the material contains equal concentrations of Mg and H, and is thus electrically inactive; however, the binding energy of H to Mg, and the migra-

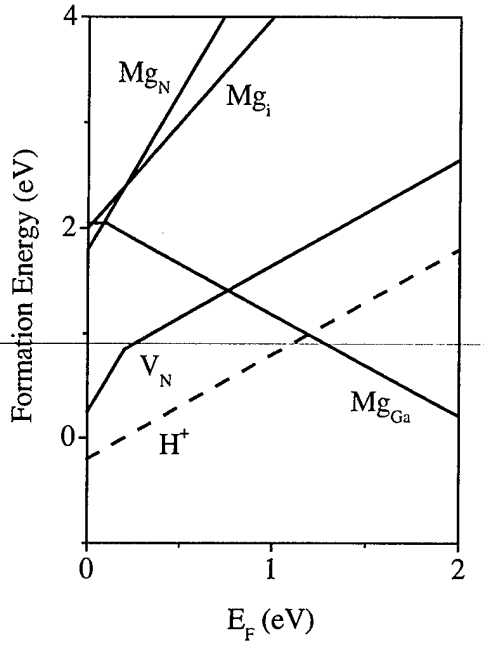


Fig. 3 Formation energies as a function of Fermi level for defects and impurities relevant for p-type GaN, under Ga-rich conditions. $E_F = 0$ corresponds to the top of the valence band.

tion barrier for H diffusion are low enough [14] to allow hydrogen to be removed from the vicinity of the acceptors by a thermal annealing procedure. Such an activation anneal is indeed known to render the Mg acceptors electrically active [12].

The incorporation of hydrogen along with Mg during the growth of p-type GaN can be regarded as a successful example of *impurity engineering*: the introduction of this additional impurity indeed enhances incorporation of the desired impurity, and leads to improved quality due to reduced compensation. This application of *codoping* is successful thanks to the relative ease with which hydrogen can be removed from the acceptor-doped layer after the growth [14]. Co-doping with donors that can *not* be removed from the acceptor-doped layer is unlikely to provide any benefits.

Finally we note that we have carried out extensive theoretical investigations [17] of alternative acceptors in GaN. Only Be has emerged as a potential improvement over Mg, exhibiting both a higher solubility and lower ionization energy. Unfortunately, Be suffers from potential self-compensation due to the incorporation of Be interstitials, which act as donors. We are currently engaged in investigations to determine whether this problem can be solved.

4 Results for ZnO

4.1 n-type doping, hydrogen

We have recently witnessed a resurgence of interest in ZnO as an optoelectronic and electronic material, triggered by advances in ZnO bulk crystal growth [18]. ZnO is an extremely versatile material, with important applications in varistors, piezoelectric transducers, and phosphors. The recent improvements in crystal quality have raised hopes for applications in light emitters; however, such device applications require improved control over conductivity.

Zinc oxide almost always exhibits strong n-type conductivity. Despite years of investigations, the source of this conductivity has remained controversial. Because of its prevalence, the n-type conductivity has traditionally been attributed to native defects. Recent first-principles investigations [19,20], however, have revealed that none of the native defects exhibits characteristics consistent with a high-concentration shallow donor. Only the vacancies have sufficiently low energies to form during synthesis of the material; zinc vacancies behave as deep acceptors, and oxygen vacancies as deep donors. The prevailing n-type conductivity can therefore not be attributed to native defects; it must thus be caused by impurities that are unintentionally incorporated.

The obvious candidates for donors in ZnO (Al or Ga on the Zn site, F or Cl on the O site) are unlikely to systematically occur as contaminants in the variety of growth environments that are used for ZnO. Hydrogen, however, has turned out to be a likely candidate [4]. This may seem surprising, since (as highlighted in Sect. 3.2) hydrogen behaves as an amphoteric impurity in the semiconductors in which it has previously been investigated. That means it always *compensates* the prevailing conductivity of the material, and cannot act as a *source* of conductivity. Hydrogen in ZnO, however, behaves differently, as can be seen from the calculated formation energy in Fig. 4: only the *positive* charge state is stable, causing hydrogen to behave exclusively as a donor. Stabilization of the neutral and negative charge states would require moving the Fermi level far above the conduction-band minimum. The hydrogen-induced level in the band structure is thus a resonance in the conduction band. Electrons placed in this resonance of course relax to the conduction-band minimum, where they can be bound to the donor in a hydrogenic state. Hydrogen thus behaves as a shallow donor in ZnO.

The reason for the high stability of H^+ in ZnO is the strong bond that can be formed between hydrogen and oxygen; indeed, interstitial H^+ sits at a distance of about 1 Å from an oxygen atom, in either a bond-center or anti-bonding configuration, with the bond between oxygen and one of its neighboring Zn atoms effectively broken. For the bond-center (BC) configuration the Zn atom moves outward over a distance equal to 40% of the bond length (0.8 Å), to a position slightly beyond the plane of its nearest neighbors.

We have calculated the vibrational frequencies of the O-H stretching and wagging modes (in the harmonic approximation); for H^+ at BC we find 3680 cm^{-1} (stretch) and 450 cm^{-1} (wag). Not surprisingly, the value for the stretch modes is close to that for H_2O molecules.

It is also interesting to consider complex formation between hydrogen and native defects. We pointed out

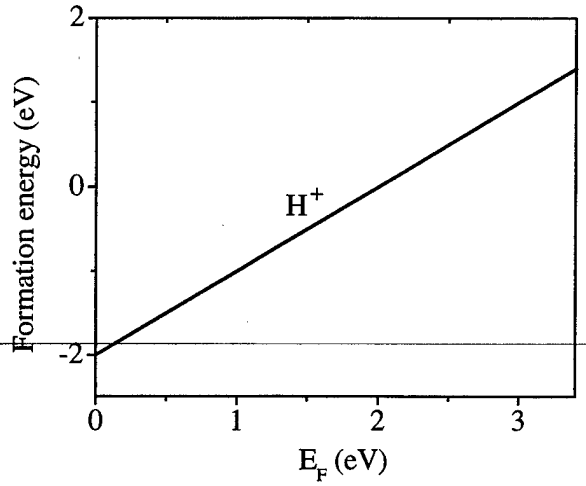


Fig. 4 Formation energy of the positive charge state of hydrogen in ZnO as a function of Fermi level. $E_F = 0$ corresponds to the top of the valence band. The chemical potential of hydrogen was chosen to correspond to H_2 molecules at T=0. Charge states other than the positive are not stable.

before that the oxygen vacancy is a *deep donor*; however, complex formation with hydrogen turns it into a *shallow donor*. The calculated binding energy, expressed with respect to H^+ and V_O^0, is 0.8 eV. The hydrogen atom is located close to the center of the vacancy (to within 0.05 Å); this configuration can thus also be regarded as a substitutional hydrogen impurity located on an oxygen lattice site.

4.2 Comparison with experiment

Experimental indications for hydrogen's behavior as a donor in ZnO were actually reported as early as the 1950s [21–23], ZnO being the first semiconductor in which the properties of hydrogen were systematically studied. Those results, however, went largely unnoticed during the upsurge in research activity on hydrogen in semiconductors that started about 30 years later. Mollwo [21] observed an increase in the conductivity of ZnO crystals exposed to hydrogen at temperatures above 200°C. The increase in the conductivity was demonstrated to be due to in-diffusion of hydrogen. An increase in conductivity upon exposure to H_2 has also been observed by Baik *et al.* [24], and by Kohiki *et al.* [25] who introduced hydrogen by proton implantation followed by annealing at 200°C. All of these experiments indicate that introducing hydrogen into ZnO does *not* result in a reduction of the conductivity, which is the expected behavior for hydrogen in other semiconductors. Instead, hydrogen shows strong behavior as a donor, consistent with our theoretical predictions.

Hydrogen is a likely candidate for an unintentional donor, since it is present in many of the growth environments commonly used for ZnO, including vapor-phase transport [18], hydrothermal growth [26], and chemical vapor deposition (MOCVD) [27]. When growth is carried out in air, water vapor can act as a source of hydrogen, and techniques such as laser ablation [28] or sputtering [29] are sometimes intentionally carried out in a hydrogen atmosphere. In addition, H_2 or H_2O may always be present as a residual gas in any high vacuum system, serving as a source of hydrogen in techniques such as molecular beam epitaxy.

Ion channeling experiments [30] may provide further evidence for the presence of interstitial H in ZnO. Ohta *et al.* [30] used ion channeling to investigate the lattice positions of the host atoms in ZnO. They found that Zn atoms have a much greater tendency to displace from the atomic rows along the $< 0001 >$ direction than oxygen atoms. This is in agreement with our result that for the lowest-energy configuration of H^+, at the bond-center site, large displacements of the Zn atoms occur. The concentration of the point defect giving rise to the displacement was estimated to be on the order of 10^{20} cm^{-3} [30]. These results are consistent with the presence of hydrogen as an unintentional donor.

Finally, we comment on the possibility of achieving p-type conductivity in ZnO. Given its size and chemical similarity to oxygen, nitrogen is probably the most promising candidate acceptor. Our preliminary investigations indicate that nitrogen would be a reasonably shallow acceptor, but with a low solubility. In addition, compensation by oxygen vacancies may occur. Co-doping with hydrogen could potentially be beneficial, as it was in the case of p-type doping of GaN. The success of this approach will depend on the ability to remove hydrogen from the p-type layer after the growth.

5 Summary

We have described a theoretical approach for studying doping of semiconductors, taking the effects of (intentional and unintentional) impurities as well as point defects into account. The framework is generally applicable, but has been applied here to identify and overcome bottlenecks in doping of wide-band-gap semiconductors. Inspection of formation energies for the relevant impurities and defects provides immediate insight in their electronic properties, as well as their expected concentrations in the material.

In GaN, we have shown that nitrogen vacancies are unlikely to be responsible for unintentional n-type conductivity; rather, impurities such as silicon or oxygen act as unintentional dopants. p-type GaN benefits from the incorporation of hydrogen during growth, which suppresses compensation and enhances the solubility of the acceptor. This is an example of *defect and impurity engineering*.

For ZnO, native defects are again unlikely as the cause of unintentionally n-type conductivity. In particular, the widely discussed oxygen vacancy is a *deep*, rather than a *shallow* donor. We have proposed hydrogen as a candidate for the unintentional donor. In contrast to other semiconductors, hydrogen is not amphoteric in ZnO, but occurs exclusively as a donor. Controlling the conductivity of ZnO thus requires careful control of hy-

drogen exposure during and after growth. We suggest that the behavior observed here for hydrogen in ZnO is likely to occur in other oxide materials as well. The fundamental insights provided by the present work should prove useful for studying and controlling hydrogen in other systems.

Acknowledgements CVdW would like to acknowledge fruitful collaborations with C. Stampfl, S. Limpijumnong, J. McCaldin, S. Pantelides, A. Kohan, G. Ceder, D. Vogel, P. Krüger, J. Pollmann, and M. Fuchs, and express his gratitude to the Fritz-Haber-Institut and Paul-Drude-Institut, Berlin, Germany, for their hospitality, and to the *Alexander von Humboldt Foundation* for a *US Senior Scientist Award.* The work described here was supported in part by the Office of Naval Research (Contract No. N00014-99-C-0161), the Air Force Office of Scientific Research (Contract No. F49620-00-C-0019).

References

1. C. G. Van de Walle, D. B. Laks, G. F. Neumark, and S. T. Pantelides, Phys. Rev. B **47** (1993) 9425.
2. J. Neugebauer and C. G. Van de Walle, Phys. Rev. B **50** (1994) 8067.
3. J. Neugebauer and C. G. Van de Walle, Appl. Phys. Lett. **68** (1996) 1829.
4. C. G. Van de Walle, Phys. Rev. Lett. **85** (2000) 1012.
5. M. Bockstedte, A. Kley, J. Neugebauer, M. Scheffler, *Comp. Phys. Commun.* **107** (1997) 187.
6. D. Vogel, P. Krüger, and J. Pollmann, Phys. Rev. B **54** (1996) 5495; Phys. Rev. B **55** (1997) 12836.
7. C. Stampfl, C. G. Van de Walle, D. Vogel, P. Krüger, and J. Pollmann, Phys. Rev. B **61** (2000) 7846.
8. F. A. Kröger, *The Chemistry of Imperfect Crystals* (North-Holland, Amsterdam 1964).
9. J. Neugebauer and C. G. Van de Walle, in *Proceedings of the 22th International Conference on the Physics of Semiconductors*, edited by D. J. Lockwood (World Scientific Publishing Co Pte Ltd., Singapore 1994), p. 2327.
10. W. Götz, R. S. Kern, C. H. Chen, H. Liu, D. A. Steigerwald, and R. M. Fletcher, Mater. Sci. Engin. B **59** (1999) 211.
11. D. P. Bour, H. F. Chung, W. Götz, L. Romano, B. S. Krusor, D. Hofstetter, S. Rudaz, C. P. Kuo, F. A Ponce, N. M. Johnson, M. G. Craford, and R. D. Bringans, Mater. Res. Soc. Symp. Proc. **449** (1997) 509.
12. S. Nakamura, N. Iwasa, M. Senoh, and T. Mukai, Jpn. J. Appl. Phys. **31** (1992) 1258.
13. J. Neugebauer and C. G. Van de Walle, Phys. Rev. Lett. **75** (1995) 4452.
14. J. Neugebauer and C. G. Van de Walle, Appl. Phys. Lett. **68** (1996) 1829.
15. N. M. Johnson and Chris G. Van de Walle, in *Hydrogen in Semiconductors II*, edited by N. H. Nickel, *Semiconductors and Semimetals*, Vol. 61, Treatise Editors R. K. Willardson and E. R. Weber (Academic Press, Boston 1999), p. 13.
16. L. Pavesi and P. Gianozzi, Phys. Rev. B **46** (1992) 4621.
17. J. Neugebauer and Chris G. Van de Walle, J. Appl. Phys. **85** (1999) 3003.
18. D. C. Look, D. C. Reynolds, J. R. Sizelove, R. L. Jones, C. W. Litton, G. Cantwell, and W. C. Harsch, Solid State Commun. **105** (1998) 399.
19. A. F. Kohan, G. Ceder, D. Morgan, and Chris G. Van de Walle, Phys. Rev. B **61** (2000) 15019.
20. C. G. Van de Walle *et al.* (unpublished).
21. E. Mollwo, Z. Physik **138** (1954) 478.
22. D. G. Thomas and J. J. Lander, J. Chem. Phys. **25** (1956) 1136.
23. J. J. Lander, J. Phys. Chem. Solids **3** (1957) 87.
24. S. J. Baik, J. H. Jang, C. H. Lee, W. Y. Cho, K. S. Lim, *Appl. Phys. Lett.* **70** (1997) 3516.
25. S. Kohiki, M. Nishitani, T. Wada, T. Hirao, *Appl. Phys. Lett.* **64** (1994) 2876.
26. M. Suscavage, M. Harris, D. Bliss, P. Yip, S.-Q. Wang, D. Schwall, L. Bouthilette, J. Bailey, M. Callahan, D. C. Look, D. C. Reynolds, R. L. Jones, and C. W. Litton, MRS Internet J. Nitride Semicond. Res. **4S1** (1999) G3.40.
27. S. Y. Myong, S. J. Baik, C. H. Lee, W. Y. Cho, and K. S. Lim, Jpn. J. Appl. Phys. **36** (1997) L1078.
28. H. Kordi Ardakani, Thin Solid Films **287** (1996) 280.
29. A. Valentini, F. Quaranta, M. Penze, and F. R. Rizzi, J. Appl. Phys. **73** (1993) 1143.
30. Y. Ohta, T. Haga, Y. Abe, *Jpn. J. Appl. Phys.* **36** (1997) L1040.

Mat. Res. Soc. Symp. Proc. Vol. 693 © 2002 Materials Research Society

Vibrational Spectroscopy of GaN:Mg Under Pressure

M.D. McCluskey, K.K. Zhuravlev, M. Kneissl,[1] W. Wong,[1] D. Treat,[1] S. Limpijumnong,[1]
C.G. Van de Walle,[1] N.M. Johnson[1]
Department of Physics and Institute for Shock Physics, Washington State University,
Pullman, WA 99164-2814, U.S.A.
[1]Xerox PARC, 3333 Coyote Hill Rd.,
Palo Alto, CA 94304, U.S.A.

ABSTRACT

The microscopic structure of Mg-H complexes in GaN has been a subject of intense theoretical and experimental investigation. In order to probe the Mg-H structure, we have studied the effect of hydrostatic pressure on the local vibrational mode (LVM) frequency. At ambient pressure, the LVM frequency is 3125 cm^{-1}, which corresponds to a N-H stretching mode. In this study, Fourier-transform spectroscopy was performed on free-standing GaN:Mg,H samples in a diamond-anvil cell, with nitrogen as a pressure-transmitting fluid. The samples had been removed from their sapphire substrate by the laser-liftoff technique. The LVM frequency was measured, at liquid helium temperatures, for pressures ranging from 0 to 5 GPa. The pressure dependence of the frequency is nonlinear: first it decreases with pressure, then it increases. Comparison with first-principles calculations allows us to derive information about the microscopic structure of the Mg-H complex. The calculated stable configuration indeed gives rise to a frequency shift consistent with experiment. Based on the comparison between theory and experiment, we can exclude the bond-center configuration, which would result in a much larger pressure derivative than experimentally observed.

INTRODUCTION

GaN is a preferred material for optoelectronic devices such as blue-violet and UV light-emitting diodes (LEDs) and lasers [1,2]. Effective *p*-type doping of GaN has been a significant challenge for the fabrication of efficient devices with long operating lifetime. Metalorganic chemical vapor deposition (MOCVD) is the dominant growth technique for III-V nitride devices, with Mg the most common *p*-type dopant. As a result of hydrogen passivation during growth, as-grown GaN:Mg is semi-insulating. It was shown empirically that low energy electron beam irradiation [3] or thermal annealing at temperatures above 600°C in an N_2 ambient [4] were required to activate the Mg acceptors. Infrared (IR) spectroscopy was used to positively identify the Mg-H complexes [5].

The stretch-mode frequency of hydrogen in the Mg-H complex was calculated to be approximately 3360 cm^{-1} [6,7], consistent with a N-H stretching mode. The LVM was observed at 3125 cm^{-1} in 4 μm thick epilayers of MOCVD-grown GaN:Mg [5]. Upon annealing, the peak at 3125 cm^{-1} decreased by a factor of two and was correlated with an increase in the conductivity. Annealed samples that are exposed to a remote deuterium plasma show a deuterium stretch mode peak at 2321 cm^{-1}. The isotopic frequency ratio is $r = \nu_H/\nu_D = 1.346$, which is very similar to that of NH_3 ($r = 1.342$), lending further support to the N-H model. The details of the microscopic structure, however, have not been resolved conclusively [8].

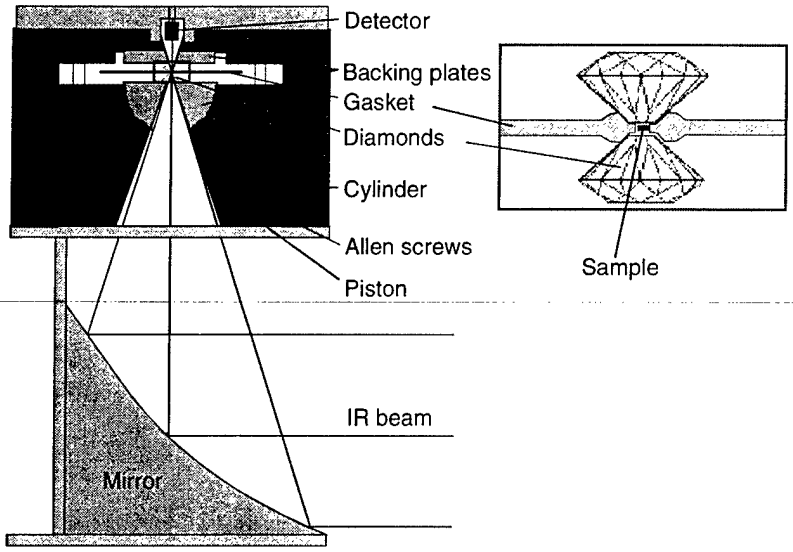


Figure 1. Cross section of DAC assembly designed for IR spectroscopy. Diamond anvils, gasket, and sample are shown in the inset.

Until recently, experimental and theoretical studies have focused on the antibonding (AB) and bond-centered (BC) configurations. Polarized IR absorption experiments by Clerjaud *et al.* [9] have indicated that the N-H bond lies at a 130° angle to the *c*-axis. This angle is incompatible with the AB configuration, in which the N-H bond occurs at ~110 ° to the *c*-axis. Recently, first-principles calculations [7,10] have shown that a novel "OA" (off-axis) configuration exhibits a 134° angle, in good agreement with Ref. [9]. Although the OA configuration is higher in energy than AB at $T = 0$, it is stabilized at high temperatures due to an increase in the entropy [10].

EXPERIMENTAL DETAILS

In order to probe the Mg-H structure, we studied the effect of hydrostatic pressure on the LVM frequency. To measure the IR spectra of semiconductors under pressure, a custom diamond anvil cell was used (Fig. 1). In this cell, force is applied by six Allen screws, which squeeze together a piston and cylinder [11]. The diamonds are supported by flat and hemispherical backing plates made from tungsten carbide. An off-axis parabolic mirror efficiently focuses the collimated IR beam onto the sample. A photoconducting Ge:Cu detector [12] is placed in close proximity to the sample, so that a large fraction of the transmitted IR light is collected. The entire assembly is placed in a Janis STVP-100 liquid-helium cryostat and kept at a temperature of 6-12 K. The low temperatures are advantageous for two reasons. First, the LVM lines sharpen with decreasing temperature. Second, the sensitive, low-noise Ge:Cu detector operates at liquid-helium temperatures.

Diamonds are transparent over a wide spectral range, although type I diamonds contain nitrogen impurities that strongly absorb from 1000 to 1500 cm^{-1}. Type II-A diamonds, used in this study, do not contain this absorption band. However, all diamonds contain a two-photon absorption band around 2100 cm^{-1}. N_2, Ar, or He are typically used as the ambient, since, unlike alcohol mixtures, they are transparent to IR light. In this work, N_2 was used. In N_2, a vibrational mode of CO_2 impurities is sensitive to pressure and was used as a precise *in situ* pressure

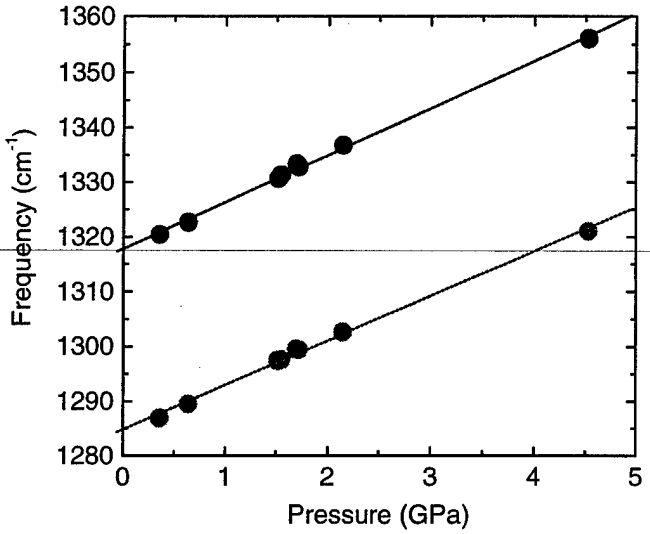


Figure 2. Pressure dependence of IR-active, two-phonon modes in GaN at a temperature of 8 K.

calibration for these IR absorption experiments [13]. This calibration has been established up to a pressure of 7 GPa.

The sample consisted of a 1 μm layer of GaN:Mg,H on a 4 μm layer of nominally undoped GaN, grown by MOCVD on a sapphire substrate. The primary difficulty in these experiments was obtaining a sufficient signal-to-noise ratio, given the very weak absorption coefficient of the Mg-H peak. To achieve that goal, the GaN epilayer was removed from the sapphire substrate by the laser lift-off technique [14]. Several (5-10) pieces were placed in the diamond-anvil cell and subjected to pressure, using liquid nitrogen as an ambient. Given the large number or samples in the DAC, the effective sample thickness was increased, thereby increasing the absorption of the Mg-H peak.

DISCUSSION

Lattice Phonons

Strong IR absorption peaks were observed in the spectral region around 1300 cm^{-1}. These peaks are attributed to two-phonon modes of the GaN lattice. The peak frequencies of two of these peaks are plotted as a function of pressure in Fig. 2. Linear fits to the data yield

$$\nu_1 = (1284.8 \pm 0.4) + (8.1 \pm 0.2)P \quad (1)$$

$$\nu_2 = (1317.8 \pm 0.4) + (8.6 \pm 0.2)P \quad (2)$$

where the frequencies ν are in units of cm^{-1} and the pressure P is in units of GPa. The pressure calibration was determined by CO_2 molecules in the solid N_2 ambient. In some experiments, however, there was not a sufficiently high concentration of CO_2 to use that method. Hence, for

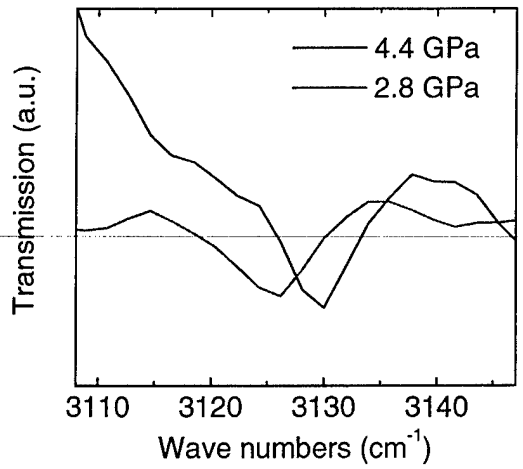


Figure 3. IR spectra of GaN:Mg,H at two different pressures, at a temperature of 8 K.

consistency, the pressure dependence of the two-phonon peaks was used as an internal pressure calibrant, via Eqs. (1) and (2).

Mg-H LVM

IR spectra of GaN:Mg,H at two different pressures are shown in Fig. 3. The LVM frequency is shown as a function of pressure in Fig. 4. The data points in Fig. 4 are a compilation of two separate experiments, in which different pieces from the same wafer were loaded into the DAC. After each spectrum was obtained, the sample was warmed to room temperature and the pressure was adjusted. The pressure was increased and decreased several times, and no hysteresis was

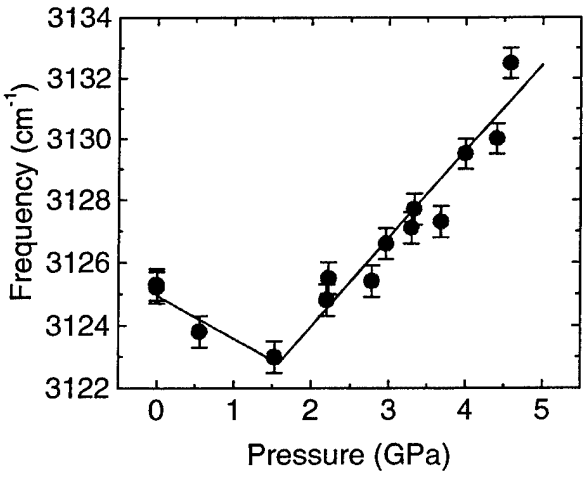


Figure 4. Pressure dependence of the Mg-H LVM frequency in GaN at a temperature of 8 K.

observed. For pressures $0<P<2$ GPa, the intensity of the peak was generally weaker than for $P>2$ GPa. Hence, in spite of many attempts, only three reliable data points were obtained for the low-pressure range. Interestingly, the LVM frequency appears to decrease with pressure, then increase. A piecewise linear fit to the data yields:

$$\nu = \begin{cases} (3125.0 \pm 0.4) - (1.4 \pm 0.4)P & P \le 1.6 \\ (3118.3 \pm 1.1) + (2.8 \pm 0.3)P & P > 1.6 \end{cases} \quad (3)$$

Theory

The experimental pressure shifts were compared with predictions from first-principles calculations, using density functional theory in the local density approximation and norm-conserving potentials with a plane-wave basis set [10]. To simulate hydrostatic pressure, the *a* and *c* lattice constants were each reduced by 1%. Assuming a bulk modulus of 200 GPa, an isotropic strain of 1% (volume decrease of 3%) corresponds to approximately 6 GPa. Given that value, the pressure derivatives of the LVM frequencies were calculated for hydrogen in the BC, AB, and OA configurations. For the BC configuration (where H was inserted into a Mg-N bond along the *c*-axis), the frequency shift was 15 cm^{-1}/GPa, much larger than the experimentally observed shifts in Eq. (3). Therefore, we can safely exclude BC as a candidate for the observed Mg-H complexes. The first-principles calculations actually show that BC is not a stable configuration: a small displacement causes it to relax to the OA configuration. The calculated AB and OA frequency shifts were –0.3 and 1.5 cm^{-1}, respectively. For the AB configuration, the N-H bond formed an angle of 109° with the *c*-axis. In the AB and OA positions, the hydrogen is not crowded by neighboring atoms, resulting in a relatively small frequency shift. The error bars for the calculated frequencies are too large to unambiguously distinguish between AB and OA based on the measured pressure shifts. However, we observe that the calculated result for OA agrees reasonably well with experiment for $P > 1.6$ GPa.

CONCLUSIONS

By measuring the pressure dependence of the GaN:Mg,H LVM and comparing with first-principles calculations, the BC configuration can be ruled out. The unresolved question is whether hydrogen is more likely to reside in the AB or OA configurations. The experimental pressure shift at larger pressures is more consistent with the OA configuration, but the error bars on the calculated frequencies do not allow an unambiguous identification. These uncertainties render it even more difficult to address the cause of the discontinuity in the slope of the Mg-H LVM vs. pressure plot. A transformation between different configurations appears unlikely, since no discontinuity is observed in the frequency. Another possibility is that the nonlinear pressure dependence is due to a "buckling" of the OA hydrogen under pressure, similar to what is observed in Si:O [15]. In future work, to test isotope effects, GaN:Mg,D samples will be investigated.

ACKNOWLEDGMENTS

This work was supported in part by NSF Grant No. DMR-9901625 and in part by AFOSR under contract # F4920-00-C-0019.

REFERENCES

1. S. Nakamura, *Mater. Res. Soc. Symposia Proc.* **482**, 1145 (1998).
2. N.M. Johnson, A.V. Nurmikko, and S.P. DenBaars, *Physics Today* **53** (10), 31 (2000).
3. H. Amano, M. Kito, K. Hiramatsu, and I. Akasaki, *Jpn. J. Appl. Phys.* **28**, L2112 (1989).
4. S. Nakamura, T. Mukai, M. Senoh, and N. Iwasa, *Jpn. J. Appl. Phys.* **31**, L139 (1992).
5. W. Götz, N.M. Johnson, D.P. Bour, M.D. McCluskey, and E.E. Haller, *Appl. Phys. Lett.* **69**, 3725 (1996).
6. J. Neugebauer and C.G. Van de Walle, *Appl. Phys. Lett.* **68**, 1829 (1996).
7. C.J. Fall, R. Jones, P.R. Briddon, and S. Öberg, *Mat.. Sci. Eng. B* **82**, 88 (2001).
8. W.R. Wampler, S.M. Myers, A.F. Wright, J.C. Barbour, C.H. Seager, and J. Han, *J. Appl. Phys.* **90**, 108 (2001).
9. B. Clerjaud, D. Côte, A. Lebkiri, C. Naud, J.M. Baranowski, K. Pakula, D. Wasik, and T. Suski, *Phys. Rev. B* **61**, 8238 (2000).
10. S. Limpijumnong, J.E. Northrup, and C.G. Van de Walle, *Phys. Rev. Lett.* **87**, 205505 (2001).
11. G. Yu. Machavariani, M.P. Pasternak, G.R. Hearne, and G. Kh. Rozenberg, *Rev. Sci. Instrum.* **69**, 1423 (1998).
12. Haller-Beeman Associates, Inc. (www.haller-beeman.com).
13. M.D. McCluskey, L. Hsu, L. Wang, and E.E. Haller, *Phys. Rev. B* **54**, 8962 (1996).
14. E.A. Stach, M. Kelsch, E.C. Nelson, W.S. Wong, T. Sands, and N.W. Cheung, *Appl. Phys. Lett.* **77**, 1819 (2000).
15. M.D. McCluskey and E.E. Haller, *Phys. Rev. B* **56**, 9520 (1997).

Mat. Res. Soc. Symp. Proc. Vol. 693 © 2002 Materials Research Society

Novel configuration of Mg-H complexes in GaN

Sukit Limpijumnong*, John E. Northrup, and Chris G. Van de Walle
Xerox PARC, 3333 Coyote Hill Road, Palo Alto, CA 94304.

ABSTRACT

We describe a novel configuration of the Mg-H complex in GaN, which is not the lowest-energy configuration at T=0 but is stabilized at elevated temperatures by the large entropy associated with a set of low-energy rotational excitations. We focus on a comparison with two types of experimental results: (1) vibrational spectroscopy using polarized light [B. Clerjaud *et al.*, Phys. Rev. B **61**, 8328 (2000)]; and (2) ion-channeling [W. Wampler *et al.*, J. Appl. Phys. **90**, 108 (2001)]. New results on the predicted quantum-mechanical delocalization of the hydrogen atom and on the effect of replacing hydrogen with deuterium are reported.

INTRODUCTION

Continued progress in nitride-based devices has been hampered by the limited carrier concentrations in p-type material. While various alternative dopants (such as Be) [1, 2, 3] are being explored, most efforts are still focused on improving the doping efficiency with Mg, the most widely used acceptor. Hydrogen is known to play an important role in p-type doping, and is abundantly present in the growth environment of many growth techniques used for GaN. Hydrogen passivates the acceptors during growth, rendering the as-grown material highly resistive. A post-growth anneal dissociates the Mg-H complexes, and removes the hydrogen from the vicinity of the acceptors, resulting in p-type conductivity. This process can be experimentally investigated by monitoring the local vibrational modes associated with the Mg-H complex. The stretching mode of this complex occurs at 3125 cm^{-1}, and disappears during the activation process [4].

Despite the established role of hydrogen in the doping process, the details of the microscopic configuration of the Mg-H complex have proved elusive. A frequency of 3125 cm^{-1} is too high to be associated with a Mg-H bond. Previous computational studies [5, 6] showed that H prefers to bond with a neighboring N atom rather than with the Mg acceptor itself. The vibrational frequencies were calculated for various configurations, and it was found that hydrogen in an antibonding position (AB_N) provided the best agreement with the experimental frequency, whereas the bond-center (BC) configuration (H inserted in the bond between Mg and N) is energetically less favorable and yields a much higher vibrational frequency (by several 100 cm^{-1}). Since then, discussions of experimental results on Mg-H complexes [7, 8] have focused on the AB_N configuration. While those experiments have helped to rule out the BC configuration, they did not confirm the AB_N configuration, and in fact raised some questions about that assignment.

Clerjaud *et al.* [7] performed multitransmission infrared spectroscopy with polarized light. They found that the N-H bond is *not* aligned along the c-axis. In fact, they found that the N-H bond (or, at least the electric dipole induced by the mode) forms an angle of 130±5° with the c-axis. This value is distinctly different from the 109° angle in the AB_N configuration.

*Permanent address: School of Physics,Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand.

Wampler *et al.* [8] performed ion channeling, using an incident analysis beam of 850 keV $^3He^+$ ions. Nuclear reaction analysis was used to detect the position of deuterium, using the $^2H(^3He,^1H)^4He$ reaction, and the proton yield was measured as a function of (small) angles between the analysis beam and *c*-axis. Wampler *et al.* performed simulations of the yield results for various possible lattice locations of the hydrogen, including $AB_{N,\perp}$, $BC_{||}$, and C (at the center of the hexagonal channel). Note that the notation "||" is used to indicate an orientation of the N-H bond parallel to the *c* axis, while $\perp$ indicates the other orientations (at ~109° with the *c* axis). None of the simulated channeling yield curves in Ref. [8] matches the r... led that the measured yield

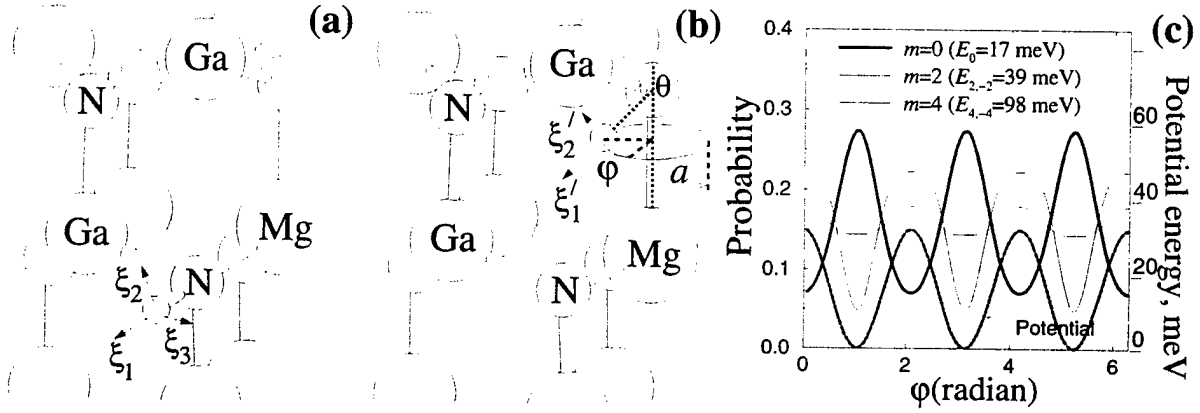


Figure 1: Perspective view of Mg-H complexes in (a) the $AB_{N,\perp}$ and (b) the $OA_{||}$ configurations. The path of H rotating around the bond axis is also shown in (b). (c) Quantum-mechanical probabilities to find H at an angle φ in the $OA_{||}$ configuration, for a few selected states (solid lines). The ground state (with energy 17 meV) is shown as a thick line, the 3rd excited state as a medium line (39 meV) and the 8th excited state as a thin line (98 meV). The potential used in the calculation is shown as a thin dashed line.

We have recently proposed a novel configuration of the Mg-H complex which is microscopically distinct from the AB_N and BC configurations. The N-H bond in this configuration, which we call $OA_{||}$ (for "off-axis"), forms an angle of 134° with *c*-axis, in very good agreement with the value 130±5 ° measured by Clerjaud *et al.* [7]. The AB_N and $OA_{||}$ configurations are illustrated in Fig. 1. The $OA_{||}$ configuration can be thought of as originating from a $BC_{||}$ configuration, but due to the large size of the Mg atom the hydrogen is pushed away from the bond axis. We find, indeed, that the BC configurations ($BC_{||}$ and $BC_\perp$) are not stable, and that a small distortion leads to a relaxation towards the $OA_{||}$ configuration.

The $OA_{||}$ configuration is not the lowest-energy configuration at T=0, but is stabilized at high temperature due to a large entropy contribution. This contribution arises from low-energy rotational excitations: in the $OA_{||}$ configuration, the H is not locked into a single position, but can rotate almost freely around the Mg-N bond axis. The free-energy calculations have been described elsewhere [9] and will not be repeated here. In this paper we focus on interpretations of experimental results in light of the novel $OA_{||}$ configuration. In particular:

(i) It is unclear whether at the low temperature (5 K) used in the vibrational spectroscopy experiments [7] the hydrogen atom is quantum-mechanically delocalized or not. We will show, however, that even if the hydrogen is delocalized the procedure used in Ref. [7] to derive the angle with the *c* axis remains valid. This will involve a generalization of Eq. (1) of Ref. [7].

(ii) The projection of the hydrogen position on the [0001] plane for the $OA_{||}$ configuration is different from any of the positions used in the channeling-yield simulations of Wampler *et al.* [8]. We will argue that the $OA_{||}$ configuration may provide a suitable match to their experimental results.

(iii) Isotopic substitution is frequently employed in experimental studies of hydrogen in semiconductors. In fact, deuterium was used in the ion-channeling studies of Ref. [8]. We will investigate the effect of isotopic substitution on the free energy of the $OA_{||}$ configuration, in order to address whether the stabilization of $OA_{||}$ also holds in the case of deuterium.

QUANTUM-MECHANICAL DELOCALIZATION

The $AB_{N,\perp}$ and $OA_{||}$ configurations are illustrated in Fig. 1. Our calculated N-H bond stretching frequency is 3092 cm^{-1} for $AB_{N,\perp}$ and 3109 cm^{-1} for $OA_{||}$. The numbers quoted here include a systematic correction, based on a comparison between calculated and experimental results for NH_3 [9]. Both of these frequencies are very close to the experimental value of 3125 cm^{-1} [4, 7] and essentially indistinguishable within our calculational accuracy. In order to differentiate between these two configurations, one needs to invoke other experiments that can probe the geometry.

The $OA_{||}$ configuration has a low energy barrier ($E_{b\varphi}$=32 meV) for H to orbit around the Mg-N bond [see Fig. 1(b)]. In the case of a free rotor ($E_{b\varphi}$=0), H would exhibit equal probability to occur at any φ, even at T=0. For the case at hand, the corrugated potential causes an uneven but symmetric (three-fold symmetry imposed by the hexagonal crystal) distribution of the probability $\mid \psi_m \mid^2$ to find H at the azimuthal coordinate φ. For simplicity, we replaced the corrugated potential with an approximation: $V_\varphi(\varphi) = E_{b\varphi}(\cos 3\varphi + 1)/2$. Since the energy barrier is low, H is delocalized and the probability is non-vanishing at all φ. In Fig. 1(c), the potential is shown with a thin-dashed line and probabilities $\mid \psi_m \mid^2$ for selected states are shown in solid lines.

DISCUSSION OF INFRARED SPECTROSCOPY RESULTS

Clerjaud *et al.* [7] used polarized infrared light to reveal additional details of the geometry of the Mg-H complex, in particular, the orientation of the electric dipole induced by the bond-stretching mode. They argued that, since hydrogen has the largest amplitude in the stretching-mode vibration, the direction of the dipole is roughly the same as the direction of the N-H bond. They found that the dipole forms an angle $\theta = 130 \pm 5°$ with the c-axis. Their conclusion was based on an analysis of the intensities of the lines obtained with different polarizations. In this analysis, it was assumed that the configuration would be static, and that H could occur in any of three equivalent configurations (imposed by the hexagonal symmetry). Here we generalize the analysis of the polarized-light measurements in Ref. [7] to include the delocalization of the hydrogen that occurs in the $OA_{||}$ configuration.

Clerjaud *et al.* wrote the intensity as a summation over three equivalent configurations:

$$I \propto \sum_{i=1}^{3} \cos^2(\mathbf{E}, \mathbf{p}_i) = \sum_{i=1}^{3} (\frac{\mathbf{E} \cdot \mathbf{p}_i}{E p_i})^2. \qquad (1)$$

$\mathbf{E}$ is the electric field of the light, and $\mathbf{p}_i$ is the electric dipole induced by the vibration of a complex in configuration i. Applying this analysis to hydrogen in the $OA_{||}$ configuration, we have to take the delocalization into account. The angle θ with the c axis remains

constant for all hydrogen locations, but hydrogen can occur at a continuum of angles φ [see Fig. 1(c)]. We define $P(\varphi)$ as the probability that hydrogen appears at an angle φ. Threefold symmetry is maintained, since the probability $P(\varphi)$ exhibits the symmetry of the crystal [see Fig. 1(c)]. The direction of the dipole is given by the unit vector $\mathbf{u}(\varphi) = [\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta]$. We can then write the generalization of Eq. (1) for the intensity of the light absorbed by the $OA_{||}$ configuration as:

$$I \propto \int_0^{2\pi} P(\varphi)(\frac{\mathbf{E}\cdot\mathbf{p}}{Ep_0})^2 d\varphi = \int_0^{2\pi} P(\varphi)(\frac{\mathbf{E}\cdot\mathbf{u}(\varphi)}{E})^2 d\varphi \qquad (2)$$

We explicitly evaluated Eq. (2) for $\mathbf{E}_{\mathrm{TE}} = E[1,0,0]$ and $\mathbf{E}_{\mathrm{TM}} = E[0,\sin\alpha,\cos\alpha]$, α defining the angle of the electric field inside the GaN crystal for the TM mode, and taking into account the symmetry of $P(\varphi)$. This results in:

$$\frac{I_{\mathrm{TE}}}{I_{\mathrm{TM}}} = \frac{\sin^2\theta}{(3\cos^2\alpha - 1)\cos^2\theta + \sin^2\alpha}. \qquad (3)$$

This is exactly the same result as reported by Clerjaud *et al.* for the "static" case. Our derivation confirms that the analysis and conclusions of Ref. [7] are equally valid in the case of delocalization of the hydrogen atom, as in the $OA_{||}$ configuration.

DISCUSSION OF ION CHANNELING RESULTS

Wampler *et al.* [8] have investigated hydrogen configurations in Mg-doped GaN with ion channeling. They used deuterium instead of hydrogen in order to enable its detection by nuclear reaction analysis (NRA), at room temperature. The angle of the analysis beam (850 keV, $^3He^+$ ions) is varied from varied from 0° up to 2° with respect to the *c*-axis. For off-axis measurements azimuthal averaging was used. The proton yield was measured as a function of incident beam angle, and simulations of the yield were performed for various configurations of H in the wurtzite lattice. Those configurations are illustrated in Fig. 2.

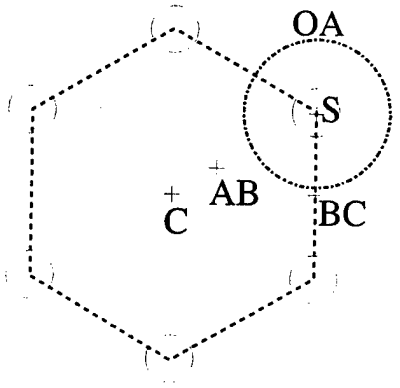


Figure 2: Lattice locations considered for the analysis of ion-channeling measurements by Wampler *et al.* [8]. Projected positions of hydrogen in the $OA_{||}$ configuration are shown as the dash-dotted circle.

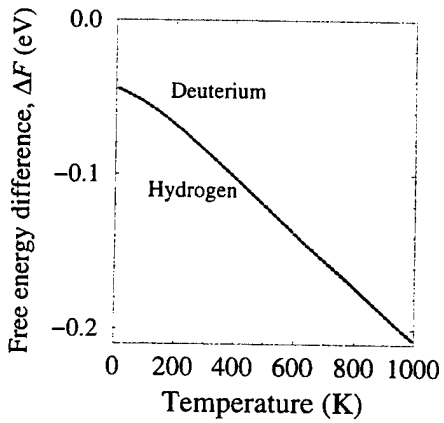


Figure 3: Free energy differences between $AB_{N,\perp}$ and $OA_{||}$ configurations of Mg-H and Mg-D complexes as a function of temperature.

The simulated yield curves (Fig. 2 of Ref. [8]) have distinctive features that can be understood by inspection of the crystal geometry. Curve C, corresponding to the center of the open channel, exhibits a narrow peak, corresponding to maximum yield when the

analysis beam is parallel to the c-axis. The C configuration could include isolated interstitial H in the neutral or negative charge states, as well as H_2 molecules. The AB curve is similar to the C curve, although the central peak is lower because H is now closer to the host atoms. This location would be appropriate for the $AB_{N,\perp}$ configuration of the Mg-H complex (or of isolated H). On the other hand, the S curve, corresponding to a location in-line with the host atoms, exhibits a broad dip, corresponding to minimum yield at 0° because the H atoms are then completely blocked by the host atoms. Once the beam is moved off-axis the H atoms are exposed. We note that the S curve would include $BC_{||}$ as well as $AB_{N,||}$ configurations. We found the former to be unstable, while the latter is about 0.3 eV higher in energy than $AB_{N,\perp}$. The BC site, finally, which corresponds to $BC_{N,\perp}$, also produces a dip, but it is narrower than in the S curve. Again, we found BC sites to be unstable in our calculations.

Experimentally, the clearest results were obtained for "gas charged" samples, which were exposed to deuterium gas at 88 kPa and 700 °C for 4 hours [8]. Other hydrogenation methods (plasma charging or implantation) carry the risk of introducing damage and exhibit a higher degree of hydrogen incorporation in configurations other than Mg-H complexes. The measured yield for the gas-charged sample exhibits a narrow peak at the center (comparable to the AB simulation) plus a broad dip, comparable to the S simulation. In order to fit this yield, Wampler *et al.* had to assume that 20% of the H was in AB sites, 30% in S sites, and 50% was randomly placed. They pointed out that this result seems to conflict with the expectation that most of the hydrogen should be in the lowest-energy configuration, which was thought to be $AB_{N,\perp}$ for the Mg-H complex. They suggested that hydrogen might be excited into higher energy configurations, possibly due to hot electrons produced by the channeling analysis ion beam.

Here we propose that the features observed in the measured yield curve can be explained by invoking the $OA_{||}$ configuration of Mg-H complexes. At room temperature, H in $OA_{||}$ is orbiting around Mg-N bond (along the c axis), with almost equal probability of occurring at any azimuthal angle. The projection of the hydrogen path is shown in Fig. 2. The nature of this configuration leads us to expect that the features it would produce in a yield curve would be some combination of the AB, S, and BC curves. We also point out that, in addition to the quantum-mechanical delocalization associated with the angle φ, H at $OA_{||}$ also exhibits a sizable vibrational amplitude due to a relatively low-frequency bending mode [coordinate ξ_2' in Fig. 1(b)]. In addition to this classical motion, H may exhibit some quantum-mechanical delocalization in the direction towards the Mg-N bond as well: indeed, while our calculations show $BC_{||}$ to be unstable, this configuration is only about 0.1 eV higher in energy than $OA_{||}$, raising the possibility of tunneling through the bond. All these contributions enhance the "S like" appearance of the yield for $OA_{||}$. We also note that the quantum-mechanical delocalization of hydrogen (as opposed to being located at discrete sites) may enhance the appearance of being located in random sites, as was found to be necessary for 50 % of the hydrogen in the simulations of Wampler *et al.* [8]. An explicit simulation of the yield curve for the $OA_{||}$ configuration would of course provide a firmer confirmation of our proposal.

DIFFERENCES BETWEEN HYDROGEN AND DEUTERIUM

Deuterium is often used in experimental studies of hydrogen in semiconductors. Using deuterium helps to overcome the sensitivity problems associated with hydrogen in secondary ion mass spectroscopy (SIMS). Vibrational spectroscopy studies often use D to check for isotope shifts, which can help in identification of local vibrational modes. And, as we saw above, the ion channeling studies of Wampler *et al.* [8] were carried out with

deuterium to take advantage of the ^{2}H(^{3}He,^{1}H)^{4}He reaction.

The $OA_{||}$ configuration of the Mg-H complex is stabilized at elevated temperatures due to the large entropy associated with low-frequency rotational excitations. These excitations are sensitive to the mass of the hydrogen atom. It is therefore important to investigate the differences that occur upon replacing hydrogen with deuterium. We examined the difference in free energies ΔF due to vibrational and rotational excitations: $\Delta F = F' - F$, where F' and F are the free energies of $OA_{||}$ and $AB_{N,\perp}$ configurations, respectively. The free energies were evaluated following the calculational approach outlined in Ref. [9].

The individual components of the free energy (corresponding to stretching, bending, and rotational excitations) incur sizable changes when we replace H by D. It turns out, however, that those changes largely cancel, producing only a small change in ΔF. As can be seen in Fig. 3, ΔF for deuterium is very similar to that of hydrogen, particularly in the temperature range of interest for stabilization of the $OA_{||}$ configuration. For all the temperatures of interest, the ΔF values for H and D differ by less than 0.01 eV. Replacing H with D therefore does not affect our conclusions about the stabilization of the $OA_{||}$ configuration at elevated temperatures.

CONCLUSIONS

We have discussed a novel configuration for the Mg-H complex in GaN ($OA_{||}$) that is consistent with the recent polarized infrared spectroscopy results of Clerjaud *et al.* [7]. In the $OA_{||}$ configuration the N-H bond forms a 134° angle with the *c*-axis, in good agreement with the 130±5° measured in Ref. [7]. We have shown that the quantum-mechanical delocalization of the hydrogen does not affect the the analysis leading to the conclusions of Ref. [7]. We also discussed the ion-channeling experiments of Wampler *et al.* [8], pointing out that the $OA_{||}$ configuration may provide a better match to the measured yields. Finally, we verified that the entropy-driven mechanisms that stabilize the $OA_{||}$ configuration are only very weakly affected by an isotopic substitution (deuterium replacing hydrogen).

Acknowledgments: This work was supported by the Air Force Office of Scientific Research, Contract No. F4920-00-C-0019, monitored by G. Witt. We thank N. M. Johnson, J. Neugebauer, and M. D. McCluskey for useful discussions and suggestions.

REFERENCES

1. J. Neugebauer and C. G. Van de Walle, J. Appl. Phys. **85**, 3003 (1999).
2. C. G. Van de Walle, S. Limpijumnong, and J. Neugebauer, Phys. Rev. B **63**, 245205 (2001).
3. J. E. Northrup, Appl. Phys. Lett. **78**, 2855 (2001).
4. W. Götz, N. M. Johnson, D. P. Bour, M. D. McCluskey, and E. E. Haller, Appl. Phys. Lett. **69**, 3725 (1996).
5. J. Neugebauer and C. G. Van de Walle, Phys. Rev. Lett. **75**, 4452 (1995).
6. A. F. Wright, Phys. Rev. B **60**, 5101 (1999).
7. B. Clerjaud, D. Côte, A. Lebkiri, C. Naud, J. M. Baranowski, K. Pakula, D. Wasik, and T. Suski, Phys. Rev. B **61**, 8328 (2000).
8. W. R. Wampler, S. M. Myers, A. F. Wright, J. C. Barbour, C. H. Seager, and J. Han, J. Appl. Phys. **90**, 108 (2001).
9. S. Limpijumnong, C. G. Van de Walle, and J. E. Northrup, Phys. Rev. Lett. **87**, 205505 (2001).

First-principles studies of beryllium doping of GaN

Chris G. Van de Walle* and Sukit Limpijumnong
Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, California 94304

Jörg Neugebauer
Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14 195 Berlin-Dahlem, Germany
(Received 12 October 2000; published 8 June 2001)

The structural and electronic properties of beryllium substitutional acceptors and interstitial donors in GaN are investigated using first-principles calculations based on pseudopotentials and density-functional theory. In *p*-type GaN, Be interstitials, which act as donors, have formation energies comparable to that of substitutional Be on the Ga site, which is an acceptor. In thermodynamic equilibrium, incorporation of Be interstitials will therefore result in severe compensation. To investigate the kinetics of Be interstitial incorporation and outdiffusion we have explored the total-energy surface. The diffusivity of Be interstitials is highly anisotropic, with a migration barrier in planes perpendicular to the *c* axis of 1.2 eV, while the barrier for motion along the *c* axis is 2.9 eV. We have also studied complex formation between interstitial donors and substitutional acceptors, and between hydrogen and substitutional beryllium. The results for wurtzite GaN are compared with those for the zinc-blende phase. Consequences for *p*-type doping using Be acceptors are discussed.

DOI: 10.1103/PhysRevB.63.245205 PACS number(s): 61.72.−y, 66.30.Jt, 71.55.Eq, 72.80.Ey

I. INTRODUCTION

Nitride semiconductors are currently being used in a wide variety of electronic and optoelectronic devices. Controllable doping is essential for all of these devices. For *n*-type doping of nitrides, a number of elements (e.g., Si) can be successfully used as dopants, and carrier concentrations exceeding 5×10^{20} cm^{-3} can routinely be achieved.[1] The situation is less favorable for *p*-type doping. Magnesium is the acceptor of choice; it can be incorporated in concentrations up to about 10^{20} cm^{-3}, but because of its large ionization energy (210 meV, Ref. 1) the resulting room-temperature hole concentration is only about 10^{18} cm^{-3}, i.e., only about 1% of Mg atoms are ionized at room temperature. Increasing the Mg concentration beyond 10^{20} cm^{-3} leads to a saturation and decrease in the hole concentration.[2] In previous work, we have proposed that the Mg solubility limit is the main cause of this behavior.[3] The limited conductivity of *p*-type doped layers constitutes an impediment for progress in device applications.

It would be desirable to identify an alternative acceptor that would exhibit higher solubility and/or lower ionization energy. Previous computational studies[4–6] have addressed a variety of candidate acceptors, including Li, Na, K, Be, Zn, Ca, and Cd. Only Be emerged as a viable acceptor, exhibiting higher solubility and lower ionization energy than Mg. However, it also emerged that self-compensation may occur due to incorporation of Be on interstitial sites, where it acts as a donor.[5] The conclusions of the previous theoretical work seem promising enough, however, to warrant a more comprehensive investigation, which is the subject of the present paper.

Experimentally, Be doping of GaN in MBE (molecular beam epitaxy) has been reported by various groups.[7–15] These experimental studies have involved both wurtzite and cubic phases of GaN; the cubic phase can be obtained by growth on cubic substrates, under appropriately tailored conditions. Another technique for Be incorporation that has been attempted is ion implantation.[16,17] No electrical conductivity results were reported; characterization was mostly by optical spectroscopy. A photoluminescence (PL) peak just below the band edge of GaN was found and attributed to Be. Assuming that this PL line results from a Be substitutional acceptor, an acceptor ionization energy can be extracted—but this has resulted in a wide range of values, ranging from 90 to 250 meV, depending on the assumptions made in the analysis. Brandt *et al.* reported that using oxygen as a codopant along with Be results in high *p*-type conductivity.[18,19] Yamamoto and Katayama-Yoshida[20] proposed an expanation for these codoping results in terms of Be_{Ga}-O_N-Be_{Ga} complexes.

The purpose of the present study is to investigate the behavior of Be in GaN in detail, with particular focus on the diffusion properties of interstitial Be. Indeed, a thorough understanding of the stability and diffusivity of interstitial Be will allow us to assess the likelihood that interstitial Be will be incorporated during growth—and if it is, whether it might be removed during a subsequent annealing procedure. To address this issue, we have mapped out the total-energy surface for an interstitial Be atom moving through the GaN lattice; this enables us to extract the barriers for Be diffusion. In addition, we have studied the interaction between interstitial Be donors (Be_{int}) and substitutional Be acceptors (Be_{Ga}). Coulomb attraction leads to formation of a complex with a large binding energy (1.35 eV). Because hydrogen is known to play an important role in *p*-type doping of GaN,[21–23] we have also investigated its interaction with Be acceptors, leading to the formation of Be-H complexes with a binding energy of 1.81 eV.

We have also performed additional investigations in order to make contact with the above-mentioned experimental studies. Because the experimental studies have involved both wurtzite and cubic phases of GaN, we have performed most of our calculations for both phases. The PL results prompted us to examine the optical ionization energy of the Be accep-

 ©2001 The American Physical Society

tor in detail, and we found a large Franck-Condon shift due to an unusual atomic relaxation of the acceptor in its neutral charge state. Because of the interest in codoping with oxygen, finally, we have performed comprehensive calculations for Be_{Ga}-O_N and Be_{Ga}-O_N-Be_{Ga} complexes.

The organization of this paper is as follows. In Sec. II, we discuss the computational approach and methods. Section III lists our results for isolated impurities, and Sec. IV contains results for complexes. Section V, finally, contains a discussion and comparison with existing experiments. Section VI summarizes the paper and contains suggestions for future experimental work.

II. THEORETICAL APPROACH

A. Pseudopotential density-functional theory

We use the density-functional theory[24] in the local density approximation (LDA) and *ab initio* norm-conserving pseudopotentials, with a plane-wave basis set.[25] Since the Ga $3d$ electrons play an important role for the chemical bonding in GaN, they cannot simply be treated as core electrons. Explicitly taking Ga $3d$ as valence electrons is one way to obtain more realistic results, as discussed in detail in Ref. 26. However, the localized nature of the Ga $3d$ states significantly increases the computational demand, requiring an energy cutoff of at least 60 Ry. We have therefore used the so-called "nonlinear core correction,"[27] with an energy cutoff of 40 Ry. We have tested this approximation by comparing formation energies for interstitial Be with results obtained using explicit inclusion of Ga $3d$ electrons as valence electrons; the values agree to within 0.1 eV.

B. Supercells

The wurtzite (WZ) structure is the lowest-energy structure of GaN, but the zinc-blende (ZB) structure is only slightly higher in energy (≈ 10 meV per two-atom unit; Ref. 28) and can be obtained by epitaxial growth on a zinc-blende substrate. We have therefore performed calculations for Be in both the zinc-blende and the wurtzite structures. As we will see, results for substitutional Be are very similar in both phases, but the properties of interstitial Be are quite different.

All impurity calculations are carried out at the theoretical lattice constant, in order to avoid any spurious relaxations. For wurtzite GaN, we find $a^{th}=3.089$ Å (compared with $a^{expt}=3.19$ Å). The calculated c/a ratio is 1.633 (experiment 1.627), very close to the ideal c/a ratio of $\sqrt{8/3}$. For zinc-blende GaN, we find $a^{th}=4.370$ Å, which is (to within 0.001 Å) $\sqrt{2}$ larger than the wurtzite lattice constant.

In order to study the atomic and electronic structure of an impurity in the GaN crystal, we construct an artificial unit cell (supercell) composed of several primitive GaN unit cells and containing one impurity. The larger the supercell size, the closer our results will be to the case of a single, isolated impurity, because interactions between impurities in neighboring supercells are suppressed. The computation time, however, increases as $N^2 \ln N$ where N is the number of atoms in the supercell. It is therefore imperative to choose the smallest possible cell size that produces reliable results for the problem under investigation. We have performed calculations for several different supercell sizes, in the process checking convergence.

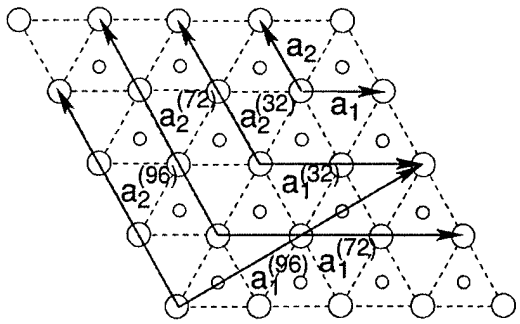


FIG. 1. Top view (along [0001] direction) of the GaN wurtzite structure: small circles represent nitrogen, large circles gallium. The shaded area corresponds to the primitive unit cell. The translation vectors for the primitive unit cell and for the 32-, 72-, and 96-atom supercells are also shown.

For the wurtzite structure, we have employed 32-, 72-, and 96-atom supercells. The 32-atom supercell is composed of eight wurtzite GaN primitive unit cells (each containing four atoms), such that each translation vector of the supercell is doubled from that of the basic unit cell. If we choose the lattice translation vectors for wurtzite GaN to be $\mathbf{a}_1=(a,0,0)$, $\mathbf{a}_2=(-a/2,\sqrt{3}a/2,0)$, and $\mathbf{a}_3=(0,0,c)$ (see Fig. 1), the translation vectors for the 32-atom supercell are defined as $\mathbf{a}_i^{(32)}=2\mathbf{a}_i$, where $i=1,2,3$. For the 72-atom supercell, the translation vectors are $\mathbf{a}_1^{(72)}=3\mathbf{a}_1$, $\mathbf{a}_2^{(72)}=3\mathbf{a}_2$, and $\mathbf{a}_3^{(72)}=2\mathbf{a}_1$. As can be noted from Fig. 1, both 32- and 72-atom supercells suffer from the problem that the separation between impurities in neighboring supercells (imagined to be located at the center of the supercell) is quite different when measured along different directions. We have therefore considered a 96-atom supercell that has translation vectors that are mutually perpendicular, leading to a cell with orthorhombic symmetry. The translation vectors of this 96-atom supercell are $\mathbf{a}_1^{(96)}=4\mathbf{a}_1+2\mathbf{a}_2$, $\mathbf{a}_2^{(96)}=3\mathbf{a}_2$, and $\mathbf{a}_3^{(96)}=2\mathbf{a}_3$.

For the zinc-blende structure, we have used 32- and 64-atom supercells, which have been extensively discussed in the context of point-defect calculations in zinc-blende semiconductors. The 32-atom supercell has bcc symmetry, while the 64-atom cell is cubic, consisting of the conventional eight-atom cubic unit cell of the zinc-blende structure doubled in each direction.

Within the supercell, we allow relaxation of several shells of host atoms around the impurity. For wurtzite, 24 host atoms were relaxed in the 32-atom cells. In the 96-atom cell, at least 46 host atoms (plus the impurity) were allowed to relax, including all atoms within 4.8 Å of any impurity. For a substitutional impurity, this corresponds to seven shells of atoms. In the zinc-blende 64-atom cell, the same relaxation radius leads to at least 44 atoms being relaxed (five shells).

Results of our convergence tests will be discussed in Sec. III A. The tests indicate that, for zinc blende, 32-atom and 64-atom supercells yield very similar results, indicating convergence. For wurtzite, the *absolute* values of formation energies are not yet converged in a 32-atom cell. The 96-atom

cell results are expected to be converged (for substitional impurities, these results are very close to those for the ZB 64-atom cells). However, 32-atom wurtzite cell calculations are very useful to obtain energy differences between different positions of the impurity in the lattice; we found these energy differences to be within 0.3 eV of those calculated in the 96-atom cell. Since mapping out the total-energy surface describing the motion of interstitial Be through the lattice requires a large number of calculations, we have used the 32-atom cell for that purpose.

C. Special k points

Brillouin-zone integrations were carried out following the Monkhorst-Pack scheme[29] with a regularly spaced mesh of $n\times n\times n$ points in the reciprocal unit cell shifted from the origin (to avoid picking up the Γ point as one of the sampling points). Symmetry reduces this set to a set of points in the irreducible part of the Brillouin zone. Convergence tests indicate that for zinc blende the $2\times2\times2$ sampling yields total energies that are converged to better than 0.1 eV in both 32-atom and 64-atom supercells. For the 32-atom wurtzite supercell with Be_{int}^{2+}, a $2\times2\times2$ set does not yield fully converged results; energy *differences* are obtained reliably, however. We also compared the total energies calculated using $1\times1\times1$ and $2\times2\times2$ sets. The former leads to only one irreducible **k** point while the latter leads to three irreducible **k** points. Although the absolute value of the total energy differs significantly between the two calculations, the energy *differences* between different Be_{int} locations change by less than 0.1 eV. The calculations required to explore the total-energy surface were therefore carried out using only one special **k** point. In the 96-atom wurtzite cell, finally, the $2\times2\times2$ **k**-point mesh produces fully converged results.

D. Charge states

Beryllium substituting on the Ga site in GaN is a single acceptor; we therefore study two charge states, Be_{Ga}^{-} and Be_{Ga}^{0}. In the case of a shallow acceptor such as Be_{Ga} the level introduced in the Kohn-Sham band structure due to the presence of the impurity is merely a perturbation of the host band structure. This "acceptor level" therefore exhibits essentially the same dispersion as the uppermost valence band. In the negative charge state, the acceptor level is filled—but in the neutral charge state, one electron is removed from this level. For a true, isolated acceptor (corresponding to a calculation in a very large supercell), the electron would be removed from the top of the valence band, at the Γ point. But in our finite-size supercells the electron is actually taken out of the highest occupied Kohn-Sham level *at the special* **k** *points*, where the band energy is lower than at the Γ point. A correction is therefore needed, obtained from the energy differences between the highest occupied state at the Γ point and the special **k** points. The magnitude of this correction ranges from $\sim$0.2 eV in the 96-atom cell to $\sim$0.5 eV in the 32-atom cells. This correction term, which we call E_{corr} and which has been defined as a positive number, has been taken into account in all our results for neutral acceptors.[30]

As we will see in the next section, the formation energy of non-neutral impurities takes into account that electrons are exchanged with the Fermi level. The Fermi level E_F is referenced to the valence-band maximum in the bulk, i.e., $E_F=0$ at the top of the valence band (E_v) in bulk GaN. We obtain E_v by calculating the position of the valence-band maximum in GaN with respect to the average electrostatic potential. We need to bear in mind, however, that the average electrostatic potential in our defect supercells is not the same as that in bulk GaN; indeed, our first-principles calculations for periodic structures (which fill all space) do not provide an absolute reference for the electrostatic potential, due to the long-range nature of the Coulomb potential. An explicit alignment procedure is thus needed between the electrostatic potential in the defect supercell and the electrostatic potential in the bulk. This problem is similar to that of calculating heterojunction band offsets, and similar techniques can be used to address these issues.[31] Here we have chosen to align the electrostatic potentials by inspecting the potential in the supercell far from the impurity and aligning it with the electrostatic potential in bulk GaN. This leads to a shift in the reference level, ΔV, which needs to be added to E_v in order to obtain the correct alignment. The resulting shifts are taken into account in our expressions for formation energies in the next section.

Interstitial Be atoms in GaN are double donors. In this work we have focused on the +2 charge state (Be_{int}^{2+}), but we make some comments about other charge states in Sec. III C 2.

Finally, we have also studied the (Be_{Ga}-Be_{int}) complex; from the charge states of the interstitial and substitutional species, it can be inferred that this complex has a charge state of +1. The neutral charge state is discussed in Sec. IV A.

E. Formation energies

The formation energy E^f determines the concentration c of the impurity in the semiconductor, through the expression

$$c=N_{\text{sites}}\exp(-E^f/kT), \tag{1}$$

where N_{sites} is the number of sites in the lattice (per unit volume) where the impurity can be incorporated, k is Boltzmann's constant, and T is the temperature. Equation (1) in principle holds only in thermodynamic equilibrium. Growth of a semiconductor is obviously a nonequilibrium process; however, many growth environments involve high enough temperatures to ensure reasonable mobility of impurities and defects at or near the surface, establishing conditions that approximate equilibrium. Furthermore, even in nonequilibrium conditions the formation energy of an impurity (or defect) is a useful concept, since configurations with high formation energies will obviously be difficult to incorporate (or will attempt to evolve to a lower-energy state).

The formation energies of Be_{int}^{2+}, Be_{Ga}^{0}, Be_{Ga}^{-}, and $(Be_{Ga}\text{-}Be_{int})^{+}$ are defined as

$$E^f[\mathrm{Be}_{int}^{2+}] = E_{tot}[\mathrm{Be}_{int}^{2+}] - E_{tot}[\mathrm{GaN, bulk}] - \mu_{\mathrm{Be}} + 2[E_F + E_v + \Delta V(\mathrm{Be}_{int})], \quad (2)$$

$$E^f[\mathrm{Be}_{\mathrm{Ga}}^{0}] = E_{tot}[\mathrm{Be}_{\mathrm{Ga}}^{0}] - E_{tot}[\mathrm{GaN, bulk}] - \mu_{\mathrm{Be}} + \mu_{\mathrm{Ga}} - E_{corr}, \quad (3)$$

$$E^f[\mathrm{Be}_{\mathrm{Ga}}^{-}] = E_{tot}[\mathrm{Be}_{\mathrm{Ga}}^{-}] - E_{tot}[\mathrm{GaN, bulk}] - \mu_{\mathrm{Be}} + \mu_{\mathrm{Ga}} - [E_F + E_v + \Delta V(\mathrm{Be}_{\mathrm{Ga}})], \quad (4)$$

$$E^f[(\mathrm{Be}_{\mathrm{Ga}}\text{-}\mathrm{Be}_{int})^{+}] = E_{tot}[(\mathrm{Be}_{\mathrm{Ga}}\text{-}\mathrm{Be}_{int})^{+}] - E_{tot}[\mathrm{GaN, bulk}] - 2\mu_{\mathrm{Be}} + \mu_{\mathrm{Ga}} + [E_F + E_v + \Delta V(\mathrm{Be}_{\mathrm{Ga}}\text{-}\mathrm{Be}_{int})]. \quad (5)$$

$E_{tot}[X]$ is the total energy derived from a supercell calculation with one impurity X in the cell, and $E_{tot}[\mathrm{GaN, bulk}]$ is the total energy for the equivalent supercell containing only bulk GaN. μ_{Ga} and μ_{Be} are the chemical potentials of Ga and Be, respectively. E_F is the Fermi level, referenced to the valence-band maximum in the bulk. Due to the choice of this reference, we need to explicitly include the energy of the bulk valence-band maximum, E_v, in our expressions for formation energies of charged states. As discussed in Sec. II D, we also need to add a correction term ΔV to align the reference potential in our defect supercell with that in the bulk. Finally, the correction term E_{corr} that appears in the formation energy of $\mathrm{Be}_{\mathrm{Ga}}^{0}$ was discussed in Sec. II D.

The chemical potentials depend on the experimental growth conditions, which can be either Ga rich or N rich. For the Ga-rich case, we use $\mu_{\mathrm{Ga}} = \mu_{\mathrm{Ga[bulk]}}$ to place an upper limit on μ_{Ga}; this places a lower limit on μ_{N}, calculated from

$$\mu_{\mathrm{N}}^{min} = E_{tot}[\mathrm{GaN}] - \mu_{\mathrm{Ga}}, \quad (6)$$

where $E_{tot}[\mathrm{GaN}]$ is the total energy of a two-atom unit of bulk GaN, calculated for the structurally optimized wurtzite structure. For the N-rich case, the upper limit on μ_{N} is given by $\mu_{\mathrm{N}} = \mu_{\mathrm{N[N_2]}}$, i.e., the energy of N in a N_2 molecule at $T=0$; this yields a lower limit on μ_{Ga} calculated from

$$\mu_{\mathrm{Ga}}^{min} = E_{tot}[\mathrm{GaN}] - \mu_{\mathrm{N}}. \quad (7)$$

The total energy of GaN can also be expressed as

$$E_{tot}[\mathrm{GaN}] = \mu_{\mathrm{Ga[bulk]}} + \mu_{\mathrm{N[N_2]}} + \Delta H_f[\mathrm{GaN}], \quad (8)$$

where $\Delta H_f[\mathrm{GaN}]$ is the enthalpy of formation, which is negative for a stable compound. Our calculated value for $\Delta H_f[\mathrm{GaN}]$ is -0.48 eV (expt. -1.17 eV; Ref. 32). By combining Eq. (6) or Eq. (7) with Eq. (8), we observe that the host chemical potentials vary over a range corresponding to the magnitude of the enthalpy of formation of GaN.

For Be, the upper bound on the chemical potential arises from the solubility-limiting phase, which is Be_3N_2. There is no lower bound on μ_{Be}, minus infinity corresponding to the total absence of Be from the growth environment. Equilibrium with Be_3N_2 implies

$$3\mu_{\mathrm{Be}} + 2\mu_{\mathrm{N}} = 3\mu_{\mathrm{Be[bulk]}} + 2\mu_{\mathrm{N[N_2]}} + \Delta H_f[\mathrm{Be_3N_2}], \quad (9)$$

where $\Delta H_f[\mathrm{Be_3N_2}]$ is the calculated enthalpy of formation of Be_3N_2, which is -5.5 eV.[33] The experimental value of $\Delta H_f[\mathrm{Be_3N_2}]$ is -6.11 eV.[34] Equation (9) allows us to relate μ_{Be} to μ_{N}, assuming equilibrium with Be_3N_2. Combining this information with the expression for the formation energy of $\mathrm{Be}_{\mathrm{Ga}}$ [Eq. (3)], we find that the *lowest* formation energy (and hence the *highest* concentration of $\mathrm{Be}_{\mathrm{Ga}}$, i.e., the solubility limit) will occur under *nitrogen-rich* conditions, i.e., when $\mu_{\mathrm{N}} = \mu_{\mathrm{N[N_2]}}$.[35]

Ultimately, one wants to calculate concentrations of native defects or impurities. For that purpose, one would choose a particular set of atomic chemical potentials (corresponding to the growth conditions), and substitute the formation energies from Eqs. (2)–(5) in the expression for concentrations, Eq. (1). This still leaves the Fermi level as a free parameter. This remaining variable is fixed by imposing the condition of charge neutrality: the total charge resulting from charged defects or impurities, along with the charge due to any free carriers in valence and conduction bands, needs to be equal to zero. This leads to a unique position for the Fermi level.

It turns out to be very informative to explicitly plot the dependence of formation energies on Fermi levels: this immediately provides insight into the electrical activity (donor or acceptor character), and shows whether certain defects will act as compensating centers. We will therefore present our results in this format.

F. Thermal and optical ionization energies

The thermal ionization energy E_A of the $\mathrm{Be}_{\mathrm{Ga}}$ substitutional acceptor is defined as the transition energy between the neutral and negative charge states of the dopant [sometimes the notation $\epsilon(0/-)$ is also used for this quantity]. This transition energy is defined as the Fermi-level position where the formation energies of these two charge states are equal, i.e.,

$$E^f[\mathrm{Be}_{\mathrm{Ga}}^{-}](E_F = E_A) = E^f[\mathrm{Be}_{\mathrm{Ga}}^{0}]. \quad (10)$$

From Eqs. (3) and (4), it then follows that

$$E_A = E^f[\mathrm{Be}_{\mathrm{Ga}}^{-}](E_F = 0) - E^f[\mathrm{Be}_{\mathrm{Ga}}^{0}] \quad (11)$$

$$= E_{tot}[\mathrm{Be}_{\mathrm{Ga}}^{-}] - E_{tot}[\mathrm{Be}_{\mathrm{Ga}}^{0}] + E_{corr} - E_v - \Delta V(\mathrm{Be}_{\mathrm{Ga}}). \quad (12)$$

For purposes of defining the thermal ionization energy, it is implied that for each charge state the atomic structure is relaxed to its equilibrium configuration. The atomic positions in these equilibrium configurations are not necessarily the same for both charge states. Indeed, we will see that they are quite different for the neutral and negative charge states of $\mathrm{Be}_{\mathrm{Ga}}$.

This difference leads to an interesting and sizable effect when the ionization energy of the dopant is determined optically. Assume the following simplified picture of a photoluminescence experiment. The exciting light creates electron-hole pairs. The holes can be trapped at $\mathrm{Be}_{\mathrm{Ga}}^{-}$ centers,

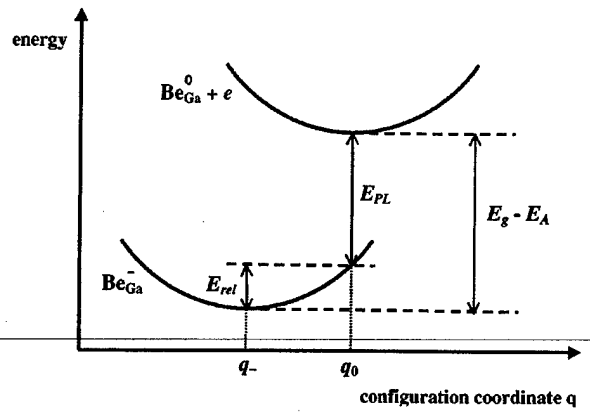
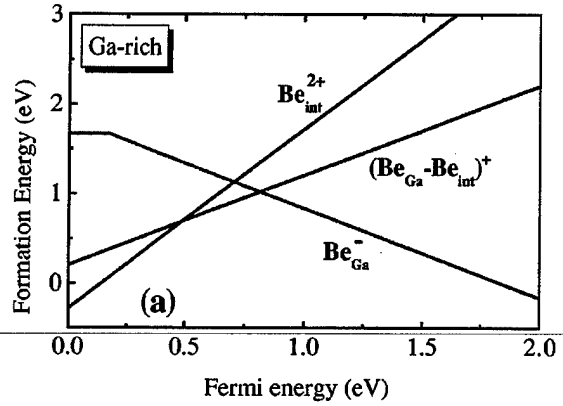


FIG. 2. Schematic configuration coordinate diagram illustrating the difference between thermal and optical ionization energies. The curve for Be^0_{Ga} is vertically displaced from that for Be^-_{Ga} assuming the presence of an electron in the conduction band. E_{rel} is the Franck-Condon shift, i.e., the relaxation energy that can be gained, in the negative charge state, by relaxing from configuration q_0 (equilibrium configuration for the neutral charge state) to configuration q_- (equilibrium configuration for the negative charge state).



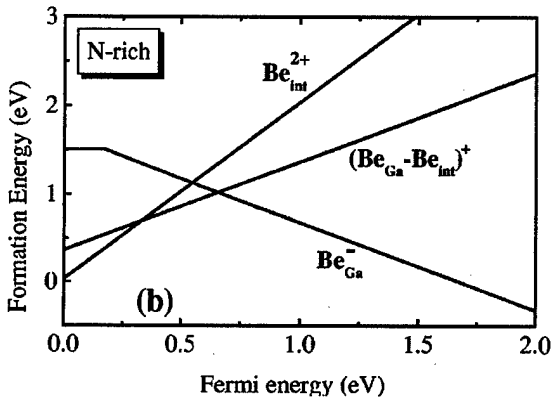


FIG. 3. Calculated defect formation energies as a function of Fermi level for various configurations of Be impurities in wurtzite GaN, under (a) Ga-rich conditions and (b) N-rich conditions.

turning them into Be^0_{Ga}. Using our definition of the thermal ionization energy E_A, the equilibrium configuration of the $Be^0_{Ga}+e$ state (where e is an electron at the bottom of the conduction band) is E_g-E_A higher than the equilibrium configuration of Be^-_{Ga}, where E_g is the band gap. Electrons in the conduction band can then recombine with the hole on the acceptor, as illustrated in Fig. 2. This leads to emission of a photon with energy E_{PL}. During this emission process, the atomic configuration of the acceptor remains fixed—i.e., in the final state, the acceptor is in the negative charge state, but with a structure (configuration coordinate q_0) that is the same as that for the neutral charge state. The difference between the energy of this configuration and that of the equilibrium configuration q_- is the relaxation energy E_{rel} (the Franck-Condon shift). From Fig. 2 it is clear that $E_{PL}=E_g-E_A-E_{rel}$. If the optical ionization energy E_A^{opt} is defined as the energy difference between the band gap and the PL line, we find that $E_A^{opt}=E_g-E_{PL}=E_A+E_{rel}$. This simplified picture ignores excitonic effects, etc., but it does show that the ionization energy extracted from an optical measurement should be larger than the thermal ionization energy E_A by an amount E_{rel}. If the atomic configuration in the two charge states is significantly different (as will turn out to be the case for Be_{Ga}), E_{rel} can be sizable.

III. RESULTS FOR ISOLATED IMPURITIES

A. The substitutional Be_{Ga} acceptor in zinc-blende and wurtzite GaN

The formation energies as a function of E_F are shown in Fig. 3 for Ga-rich and N-rich conditions. As discussed in Sec. II E, N-rich conditions favor the incorporation of Be on Ga sites, leading to the largest possible Be_{Ga} concentration (the solubility limit). Moving toward Ga-rich conditions does not alter our qualitative conclusions; numerical results for any choice of the chemical potentials can always be obtained, of course, by referring to the expressions for formation energies in Eqs. (2)–(5). In particular, by combining these expressions with the information on chemical potentials in Eqs. (8) and (9), we find that by moving from N-rich to Ga-rich conditions the formation energy of Be_{Ga} increases by $\frac{1}{3}|\Delta H_f[\mathrm{GaN}]|$, while the formation energy of Be_{int} decreases by $\frac{2}{3}|\Delta H_f[\mathrm{GaN}]|$. Consequences of these results for optimizing growth conditions will be discussed in Sec. V.

The formation energies are also summarized in Table I; the numbers given there assume the Fermi level is located at the top of the valence band. We notice that, for the substitu-

TABLE I. Calculated formation energies for Be substitutional acceptors (Be_{Ga}), Be interstitial donors (Be_{int}), and Be_{Ga}-Be_{int} complexes in wurtzite (WZ) and zinc-blende (ZB) GaN. The formation energies [see Eqs. (2)–(5)] are given for $E_F=0$, i.e., E_F located at the top of the valence band, and for nitrogen-rich conditions. The dependence of formation energies on Fermi level is shown in Fig. 3.

	Formation energies (eV)	
	WZ	ZB
Be^-_{Ga}	1.68	1.67
Be^0_{Ga}	1.51	1.47
Be^{2+}_{int}	0.04	−0.57
$(Be_{Ga}\text{-}Be_{int})^+$	0.37	−0.01

TABLE II. Relaxations around Be substitutional acceptors (Be_{Ga}) in wurtzite (WZ) and zincblende (ZB) GaN. Δd_{Be} denotes the displacement of the Be atom from its nominal lattice site, expressed as a percentage of the bulk Ga-N bond length. $\Delta d_{\parallel}$ and $\Delta d_{\perp}$ denote the percentage of change in the Be-N bond length, again expressed referenced to the bulk Ga-N bond length. For wurtzite, the symbol $\parallel$ denotes the direction parallel to the c axis ([0001]); the other bond directions are denoted by the symbol $\perp$. For zinc blende, the symmetry around the impurity is lowered to C_{3v}, with $\parallel$ denoting the particular [111] direction in which the impurity is displaced, and $\perp$ denoting the other $\langle 111 \rangle$ directions.

	Be_{Ga}^{-}		Be_{Ga}^{0}	
	WZ	ZB	WZ	ZB
Δd_{Be}	0.3%	0%	10.7%	11.4%
$\Delta d_{\parallel}$	−4.6%	−4.9%	12.2%	14.5%
$\Delta d_{\perp}$	−5.4%	−4.9%	−8.7%	−9.3%

tional acceptor Be_{Ga}, the results for wurtzite and zinc-blende structures are remarkably similar, the differences being smaller than the calculational error bars. The ionization energies, as defined in Eq. (11), are 0.17 eV for the wurtzite structure and 0.20 eV for zinc blende.

Table II summarizes our results for the atomic configuration of substitutional Be in wurtzite and zinc-blende GaN. The atomic configurations of Be_{Ga}^{-} and Be_{Ga}^{0} in wurtzite GaN are also schematically illustrated in Fig. 4. This figure, as well as all subsequent schematic representations of atomic structures, is based on calculated atomic coordinates obtained from a 96-atom supercell calculation. The substitutional Be atom is surrounded by four N neighbors. A contraction of the Be-N bond length is expected since the atomic radius of Be is smaller than that of Ga: the covalent radius of Be is 0.90 Å, while that of Ga is 1.26 Å.[36] Based on this difference, and using a bond length of 1.95 Å for bulk GaN,

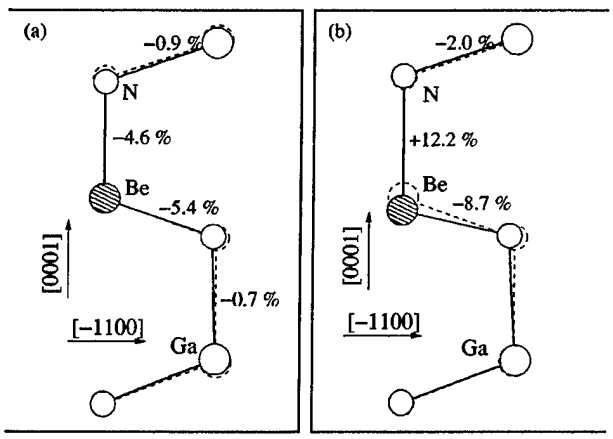


FIG. 4. Schematic representation of atomic positions in the $(11\bar{2}0)$ plane for (a) Be_{Ga}^{-} and (b) Be_{Ga}^{0} in wurtzite GaN. Large circles represent Ga atoms, medium circles N atoms, and the hatched circle represents Be. Dashed circles indicate ideal atomic positions, dashed lines bonds in the ideal lattice. The numbers denote the percentage of change in the bond lengths, referenced to the bulk Ga-N bond length (as in Table II).

we would expect a contraction of the Be-N bond length by 18% compared to the Ga-N bond length. This is indeed observed in the compound Be_3N_2, where the bond lengths are 1.50–1.64 Å, 16–23 % shorter than in GaN. Using ionic radii[37] (0.27 Å for Be^{2+} and 0.47 Å for Ga^{3+}), we would expect a contraction of about 10%. The actual magnitude of the contraction of the Be-N bond length as calculated for a Be_{Ga}^{-} impurity in GaN is less than 6%. The Be-N bond length is thus somewhat longer than is optimal for Be-N, reflecting the fact that the energy cost associated with moving the surrounding N atoms (as well as further shells of atoms) is quite large. As discussed in Ref. 5, this makes substitutional Be in GaN less energetically favorably than could be hoped based on the large Be-N bond strength. In spite of this unfavorable size match, it was observed that the formation energy of Be_{Ga} compares favorably with that of other substitutional acceptors. These observations about size mismatch do show, however, that strategies aimed at relieving the lattice strain associated with incorporating substitutional Be could have an important impact.[38]

In the *negative* charge state, the Be atom is located essentially on the substitutional lattice site. In the *neutral* charge state the situation is very different. Here the Be atom undergoes a sizable relaxation off the nominal lattice site, by more than 10% of the bond length. We also note that the $d_{\parallel}$ bond length becomes very large, indicating that the bond with one of the N neighbors is significantly weakened. Simultaneously, the bonds with the other three N neighbors contract by almost 10%.

The large off-center displacement of Be_{Ga}^{0} was found in calculations for supercells of different sizes and shapes, and occurs in both the wurtzite and the zinc-blende structures (see Table II). In the wurtzite structure, the configuration listed in Table II and depicted in Fig. 4 involves the Be atom moving along the c axis. The configuration in which Be_{Ga}^{0} moves along the direction of one of the other Be-N bonds is very close in energy (within 0.03 eV), with very similar relaxations around the Be atom. This tendency of Be to form stronger bonds with *three* of its nitrogen neighbors is not unexpected. Consider, for instance, the atomic structure of the hexagonal modification of Be_3N_2 [Ref. 39]: some of the Be atoms are surrounded by a deformed tetrahedron of N atoms, while others are surrounded by a *triangle* of nitrogen atoms. It thus seems to be energetically quite acceptable for Be to form strong bonds with only three of its four N neighbors. This tendency will also emerge when we inspect bonding configurations for interstitial Be.

Since the equilibrium configuration for the neutral acceptor is so different from that for the negative charge state, we expect the relaxation energy E_{rel} (as defined in Sec. II F and Fig. 2) to be quite sizable. E_{rel} is the energy difference, for Be_{Ga}^{-}, between the atomic configuration that yields the minimum energy for the neutral charge state (i.e., with large lattice relaxation) and the atomic configuration that yields the minimum energy for the negative charge state (i.e., with Be on the ideal lattice site); this yields $E_{rel}=0.10$ eV. Following the discussion in Sec. II F, we thus expect the ionization energy of the acceptor as determined optically to be signifi-

cantly higher than the thermal ionization energy. Consequences of this result for experimental observations will be discussed in Sec. V B.

The large off-center relaxation of the Be acceptor (at least in the neutral charge state) is reminiscent of the large lattice relaxations experienced by donor atoms in the so-called DX centers;[40] in this process, the shallow donor (d^+) is converted into a deep acceptor (DX^-). A similar mechanism for acceptors would involve placing the acceptor in a *positive* charge state (AX^+) and letting it undergo a large lattice relaxation. We have carefully examined whether such a mechanism can occur for Be, but have found no evidence for stability of a Be_{Ga}^+ configuration.

We have examined the stability of Be_{Ga} by displacing the Be atom from the substitutional site; such a displacement would be the first step in a process where the Be would move away from the substitutional site, forming a Be interstitial and leaving a Ga vacancy behind. We investigated various configurations in which a Be interstitial was placed in the vicinity of a Ga vacancy. When the overall system was placed in a negative charge state (corresponding to the Be_{Ga}^- reference state), the system immediately relaxed back, without any barrier, to the Be_{Ga} configuration. For other charge states, we found it impossible to obtain converged calculations for the system with Be moved off the substitutional site, indicating the instability of such configurations. Although it is relatively easy to form Be interstitials (see Sec. III C), the energy cost to create a Ga vacancy is prohibitively high, particularly in p-type GaN;[41] it is therefore exceedingly unlikely that substitutional Be would leave the Ga site.

B. Interstitial Be (Be_{int}) in zinc-blende GaN

The formation energy of the interstitial species Be_{int}^{2+} in zinc-blende GaN is listed in Table I. This formation energy is negative. One should keep in mind, however, that the formation energy depends on the Fermi-level position [see Eq. (2)]; the value given in Table I is for $E_F=0$, i.e., E_F positioned at the valence-band maximum. The formation energy increases (and becomes positive) when the Fermi level is higher in the band gap (see Fig. 3). As described in Sec. II E, the final Fermi-level position is determined by charge neutrality, and this position will always correspond to positive formation energies. Table I shows that the formation energy is higher in the wurtzite structure, though still low enough to be a concern for compensation, as discussed in Sec. III C and Sec. V.

The global minimum for Be_{int}^{2+} in ZB GaN is at the tetrahedral interstitial site surrounded by nitrogen atoms (T_d^N). The Be atom is essentially at the center of the cage, and the relaxation of the N atoms is such that the Be-N distance is approximately 1.73 Å; the N atoms thus move toward Be by more than 0.20 Å. Interstitial diffusion in the zinc-blende structure proceeds via a path through hexagonal rings, with an intermediate position at the tetrahedral interstitial site surrounded by Ga atoms (T_d^{Ga}). The energy difference between T_d^{Ga} and T_d^N can thus serve as an estimate for the diffusion activation energy; this value is equal to 3.74 eV. We conclude that interstitial diffusion of Be has a high barrier in zinc-blende GaN.

C. Interstitial Be (Be_{int}) in wurtzite GaN

The formation energy of the interstitial species Be_{int}^{2+} is shown in Fig. 3 and listed in Table I. Figure 3 indicates that the formation energy of the interstitial donor is low enough to be a serious concern for compensation. We already pointed out that this concern is even more severe in the zinc-blende structure, where the formation energy of the interstitial is 0.61 eV lower.

When trying to grow p-type GaN doped with Be acceptors, one should thus be aware of potential compensation by Be interstitials. The extent to which interstitials form a problem, as well as the extent to which their incorporation can be controlled, depend sensitively on their diffusivity. The fact that Be interstitials can be readily incorporated is due to the small atomic size of the Be atom; one might therefore suspect that this small atom could also diffuse readily through the lattice.

1. Total-energy surfaces

To study the behavior of the Be interstitial in GaN, we have calculated the total energy of Be_{int} at various locations in GaN. For each position of Be, the surrounding host atoms are allowed to relax. The resulting energy values as a function of the coordinates of the Be position, $\mathbf{R}_{Be}$, define the energy surface: $E=E(\mathbf{R}_{Be})$. The energy surface is therefore a function of three spatial dimensions. In order to obtain accurate results, a large data base of energy values (for a large number of spatial coordinates) is needed. In order to render the computational effort tractable, we carried out these calculations in 32-atom wurtzite cells with one k point. As discussed in Sec. II C, this affects energy differences between different Be sites by less than 0.4 eV. This accuracy is sufficient not to affect the qualitative shape of the energy surface. Local minima and saddle points identified in this investigation were subsequently also calculated using the 96-atom supercell.

We calculated 59 different locations in the irreducible portion of the wurtzite unit cell. These locations were not equally spaced in the crystal, because we chose to calculate more points near the important locations (the local minima and saddle points). For each local minimum, we also allowed Be_{int} to relax to ensure that the actual minimum energy and lattice structure were obtained.

In order to analyze and visualize the total-energy surface, we want to make optimal use of symmetry information.[42] Since the energy surface is a function of $\mathbf{R}_{Be}$, it possesses the full symmetry of the crystal. To make effective use of these symmetries, an analytic description of the surface is essential. We achieve this through expansion in a basis set with the appropriate symmetry. We then perform a fitting of the total-energy results to this set of basis functions using a least-squares method. As basis functions, we use symmetrized plane waves that reflect the full symmetry of the wurtzite crystal.[42] We also make use of the physical fact that

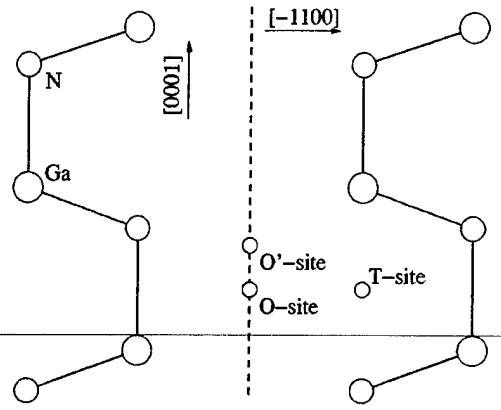


FIG. 5. Schematic representation of atomic positions in the $(11\bar{2}0)$ plane of wurtzite GaN. The large circles represent Ga atoms, medium circles N atoms. The high-symmetry interstitial sites are indicated: O is the octahedral interstitial site and T the tetrahedral interstitial site. O' indicates the position of the global minimum for Be_{int}^{2+}.

the Be impurity can never approach any host atom too closely (exchange processes, which would carry a very high energy cost, are not included in the total-energy surface). To this end, we add some *ad hoc* high-energy values near the host atoms to the calculated data set. This addition does not affect the shape of the energy surface in the relevant regions away from the atoms.

Since the energy surface is a function of three spatial dimensions, it is difficult to present the results in a single plot. For visualization purposes, we need to show a cut of the energy surface, restricting the Be coordinates to a single plane. Judicious choice of such planes ensures that we convey all the essential information, i.e., stable and metastable sites as well as barriers between them. For the system considered here, the $(11\bar{2}0)$ plane (see Fig. 5) and the (0001) plane (perpendicular to the c axis) are the relevant planes. The energy surface can then be displayed as a contour plot or as a perspective plot of the energy (along the z axis) as a function of coordinates in the plane.

In the wurtzite structure, there are two distinct types of interstitial open space, as shown in Fig 5. T is the tetrahedral interstitiitial site (or cage site). This site is equidistant from four Ga and four N atoms. O is the octahedral interstitial site (or tunnel site). This site is in the "interstitial channel" along the c axis. The O site is equidistant from six Ga and six N atoms. Both sites are obvious candidates for local minima (or, more generally, extrema) for Be_{int}.

2. Be_{int} near the O site

Placing the Be interstitial at the O site and allowing all atoms to relax results in a position (which we call the O' site) where Be optimizes its distance to its N neighbors. This position constitutes the global minimum in the total-energy surface for Be_{int}^{2+} in GaN. As can be seen in Fig. 6(a), the Be atom stays along the axis of the open channel, and moves close to the plane of the N atoms. The resulting Be-N distance (1.59 Å) is close to the optimum Be-N bond length as observed in, say, Be_3N_2. In the process, the N atoms move toward Be by 0.22 Å.

The formation energy at this global minimum is 0.61 eV higher than for Be in ZB GaN (see Table I). One might think that the cause for the higher formation energy in wurtzite is the smaller size of the interstitial cage. However, the fact that for Be_{int} at the O' site the nitrogen neighbors actually move *toward* the interstitial indicates that available volume is not an issue. The Be atom clearly likes to bind to the N atoms; in the open channel of the WZ structure, it can only effectively form bonds with *three* N atoms. At the T_d^N site in ZB, on the other hand, it can strongly bind to *four* N neighbors. We suggest that this is the reason for the lower formation energy in zinc-blende GaN. We also find that the bonding between the Be and N atoms exhibits a significant covalent component, in addition to the ionic component; this emerges from an inspection of the charge density corresponding to various eigenstates throughout the valence band.

For this global minimum, we have also examined other charge states. The Be interstitial does not introduce any Kohn-Sham states within the LDA band gap, indicating that it behaves as a shallow donor. We can calculate the formation energy of the +1 and 0 charge states by placing one or two electrons in the lowest unoccupied level. As for the case of shallow acceptors, the formation energies need to be corrected with a term E_{corr} reflecting the energy difference between electrons at the special point and at the Γ point, as discussed in Sec. II D. In addition, these formation energies suffer from the LDA band gap error; however, if we focus on energy *differences* between charge states, as appropriate for the calculation of ionization energies, then the results do not directly depend on the LDA error if we express the ionization level with respect to the conduction-band minimum.

Following this procedure, we find the +/0 transition to correspond to an effective-mass state, while the ++/+ level is slightly deeper, namely, 0.09 eV below the conduction band. We will revisit this result in Sec. V B.

Coming back to Be_{int}^{2+}, the features of the energy surface in the neighborhood of the O' site can be obtained from Fig. 7. This figure includes a contour plot [Fig. 7(a)] and a perspective plot [Fig. 7(b)] for the $(11\bar{2}0)$ plane that cuts through the hexagonal channel. We discuss these plots in more detail in the following sections.

3. Be_{int} at the T site

We place Be_{int} at the ideal T site and allow the host atoms to relax. If we allow Be_{int} to relax along [0001], we find that it stays almost at the ideal position of the T site; we therefore do not need to introduce a separate notation for the minimum-energy site, as we did in the case of the O' site. The T site is thus a local minimum for displacements parallel to [0001]. However, when we allow displacements perpendicular to [0001] (and move Be slightly off this axis, in order to break the symmetry), we find that the T site is a local maximum in the plane perpendicular to [0001]. We conclude that T is actually a saddle point. A contour plot and a perspective plot for the $(11\bar{2}0)$ plane that cuts through the T site

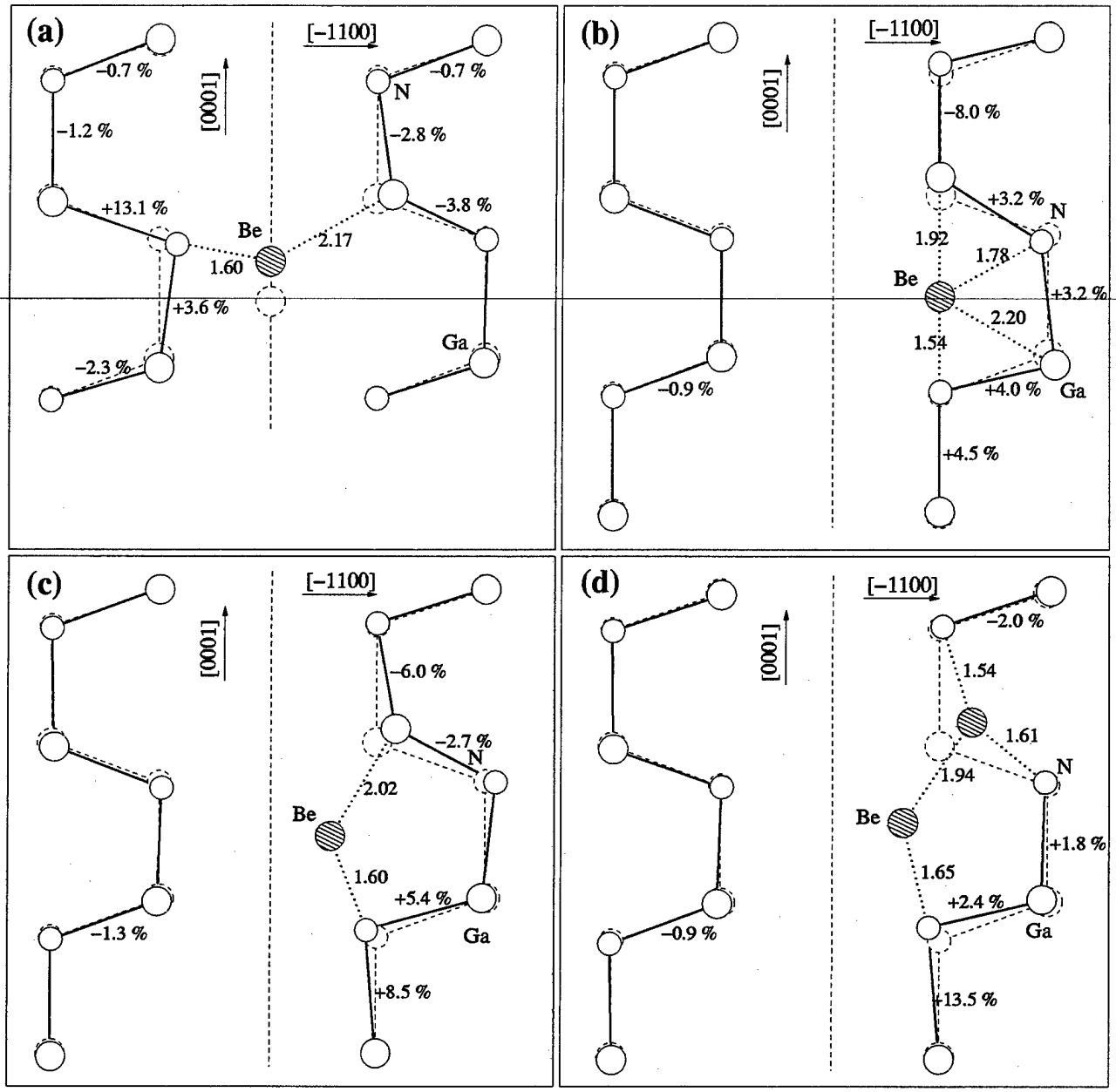


FIG. 6. Schematic representation of atomic positions in the $(11\bar{2}0)$ plane for Be_{int}^{2+} at various interstitial sites in wurtzite GaN. Large circles represent Ga atoms, medium circles N atoms, and the hatched circle represents Be. Dashed circles indicate ideal atomic positions, dashed lines bonds in the ideal lattice. Distances between Be and its neighbors are given in Å and changes in Ga-N bond lengths are given as the percentage change from the bulk Ga-N bond lengths. The panels correspond to a Be interstitial in the following configurations: (a) O' site; (b) T site; (c) A site; and (d) the Be_{Ga}-Be_{int}^{+} complex.

are shown in Figs. 7(a) and 7(b), respectively. The corresponding plots for the (0001) plane that cuts through the T site are shown in Figs. 7(c) and 7(d). Figure 7(d) clearly shows that T is a local maximum in the (0001) plane. In the $(11\bar{2}0)$ plane, the T site is a local minimum in the [0001] direction and a local maximum in the $[-1100]$ direction.

The position of the Be atom and the relaxations of the host atoms are shown in Fig. 6(b). Although the host atoms undergo quite large relaxations, the Be atom is squeezed rather tightly between the Ga and N atoms along the c axis, with a Be-Ga distance of 1.92 Å and a Be-N distance of 1.54 Å. In spite of this tight fit, the T site is only 1.18 eV higher in energy than the global minimum at the O' site (as calculated in the 96-atom supercell; in the 32-atom cell, the energy difference is 0.90 eV). The energy at T is probably lowered due to the fact that the Be atom can now bond with four N neighbors; but the size constraints, in particular the repulsion with the Ga atom along the c axis, raise the energy.

4. Be_{int} at the A site

In the process of exploring the total-energy surface for Be_{int}, we discovered an additional local minimum. Figures 7(a) and 7(b) show a contour plot and a perspective plot for

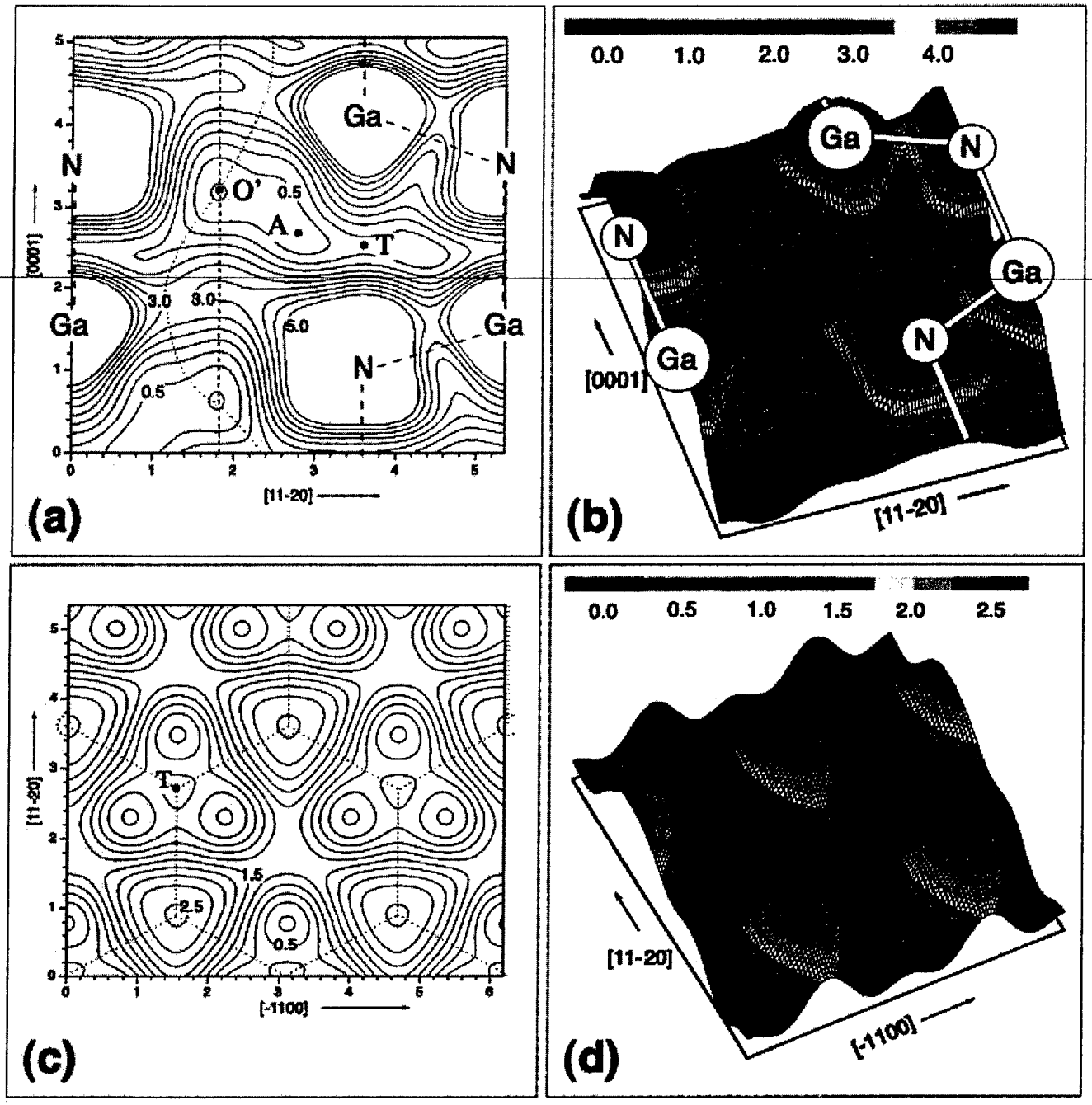


FIG. 7. (Color) Calculated total-energy surfaces for Be interstitials in GaN. (a) Contour plot and (b) perspective plot of the total-energy surface for Be_{int}^{2+} in a $(11\bar{2}0)$ plane through the host atoms (as defined in Fig. 5). In the contour plot, the interval is 0.5 eV and the real-space dimensions along both directions are in Å. (c) Contour plot and (d) perspective plot of the total-energy surface for Be_{int}^{2+} in a (0001) plane through the T site. This (0001) plane thus also cuts through the middle of Ga-N bonds along the c axis. In the contour plot, the interval is 0.25 eV and the real-space dimensions along both directions are in Å. Dashed circles show the projections of the atomic positions of the host atoms and dashed lines show the projections of Ga-N bonds.

the $(11\bar{2}0)$ plane that cuts through both T site and O site. We notice that there is a local minimum located between the T site and the O' site; we will call this site the A site. The atomic positions for Be_{int} at the A site are shown in Fig. 6(c). This local minimum has a formation energy only 0.70 eV higher than the global minimum (as calculated in the 96-atom supercell; in the 32-atom cell, the energy difference is 0.30 eV). Again, the driving force seems to be the formation of Be-N bonds; two of the N atoms move toward Be by 0.19Å, while a third moves by 0.08 Å. This results in three Be-N bonds with bond lengths of 1.60 Å. Simultaneously, a Ga atom that would prevent Be from approaching N moves

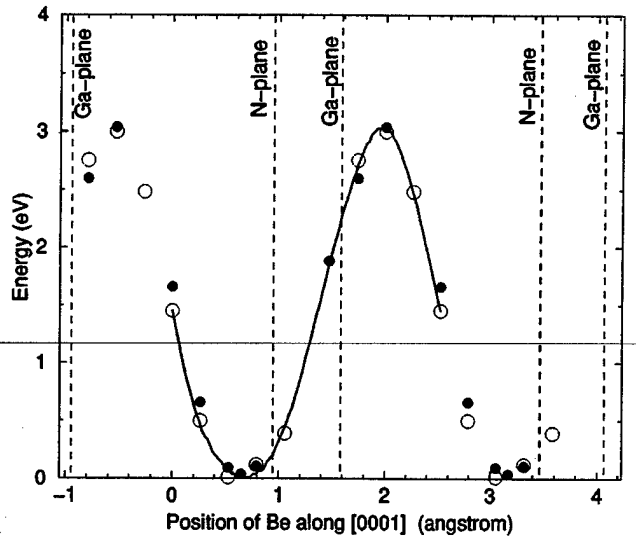


FIG. 8. Total energy of Be_{int}^{2+} at positions along the center of the open channel in WZ GaN, measured from the O site, and referenced to the energy at the global minimum (O'). Solid circles: the results from $1\times1\times1$ k-mesh calculations; open circles: $2\times2\times2$ k-mesh calculations; solid line: interpolation curve.

away, undergoing a displacement of almost 20% of the Ga-N bond length. Still, the resulting Be-Ga distance of 2.02 Å is probably the determining factor limiting the stability of this site.

5. Diffusion path for Be_{int}

The total-energy surfaces allow us to study the diffusion path of the Be interstitial along different directions. We focused on the +2 charge state, which is the relevant charge state for Be_{int} at its global minimum (see Sec. III C 2). Inspection of the electronic structure for various positions of Be along its migration path indicates that there is no tendency for localized electronic levels to move into the band gap, and hence no possibility for charge-state changes to occur.

To study the diffusion path along the [0001] direction, we first calculated the formation energies for Be_{int} at several positions along a [0001] direction through an O site (i.e., along the center of the open channel). The results are shown in Fig. 8. This diffusion path yields a diffusion barrier of 3.1 eV (in the 32-atom cell, very close to the 3.0 eV obtained in a 96-atom cell). Inspection of the full total-energy surface, as shown in Fig. 7, indicates that the lowest-energy path does not run exactly along the center of the channel. The dashed line shows the diffusion path along the center of the channel, while the dotted line shows the actual path; the barrier along the latter path is only slightly lower, at 2.90 eV.

For diffusion in a direction perpendicular to the c axis, the interstitial travels between O' sites in adjacent channels via a path that has a saddle point at the T site. The diffusion barrier is thus equal to the energy difference between the T and O' sites, which is 1.18 eV (in the 96-atom cell; 0.90 eV in the 32-atom cell).

We thus find a very large difference in barrier heights for diffusion parallel to [0001] vs perpendicular to [0001]. A rough estimate of the temperature at which the interstitials would be mobile can be obtained by taking the usual definition of an activation temperature, i.e., the temperature at which the jump rate is one per second, and assuming a prefactor of 10^{13} s^{-1}, i.e., a typical phonon frequency. With a 1.18 eV barrier, the Be interstitial should be mobile in directions perpendicular to the c axis at temperatures of about 180 °C. Diffusion along the c axis, on the other hand, with a barrier of 2.90 eV, would require temperatures of about 850 °C.

The large anisotropy of the diffusion barriers can be attributed to the features of Be bonding with the host atoms that we described above: Be_{int}^{2+} prefers to be close to at least three N atoms, with a bond length comparable to that in Be_3N_2. At the same time, Be_{int}^{2+} tries to stay as far away as possible from the Ga atoms, with which it has a repulsive interaction. As can be seen from Fig. 8, the saddle point for migration along [0001] occurs close to the plane of the Ga atoms. These atoms form part of a hexagonal ring in the "chair" configuration; in this configuration, the center of the triangle formed by the three N atoms is displaced from the center of the triangle formed by the Ga atoms by ~0.6 Å along the c axis. Because of the Be-Ga repulsion, squeezing through the triangle of Ga atoms carries a large energy cost. Simultaneously, the Be atom is far away from the N atoms with which it prefers to bond. These features conspire to create a large barrier for Be motion along the c axis.

For motion perpendicular to the c axis, the Be atom also has to squeeze through hexagonal rings, but this time these rings have the "boat" configuration. In this configuration, the center of the triangle formed by the N atoms coincides with the center of the triangle formed by the Ga atoms. This configuration thus allows the Be to stay more closely bonded to the N atoms, even while squeezing through the triangle of Ga atoms, resulting in a lower barrier.

IV. RESULTS FOR COMPLEXES

A. Complexes between interstitial Be and substitutional Be (Be_{int}-Be_{Ga}) in wurtzite GaN

Since interstitial Be is a donor (Be_{int}^{2+}), we expect it will bind with substitutional Be, which is an acceptor (Be_{Ga}^{-}). We explored various possible atomic configurations for the complex; the final stable geometry is illustrated in Fig. 6(d). We notice a similarity with the A site configuration of the Be interstitial, ilustrated in Fig. 6(c). Indeed, although this metastable site for Be_{int}^{2+} was somewhat higher in energy than the O' site, the Be interstitial is quite close to a substitutional Ga site in this A site configuration. This offers the possibility for forming a strong bond to a Be_{Ga} substituting on this site.

The relaxation can be described as follows. Be_{int} moves toward Be_{Ga} while Be_{Ga} moves away slightly in the opposite direction. The two Be atoms form a bond with a bond length of 1.94 Å. An alternative way of looking at this configuration is as a split interstitial, with the two Be atoms sharing the substitutional Ga site; the center of the Be-Be bond is displaced 0.46 Å from the ideal Ga position. Both Be atoms

TABLE III. Calculated binding energies E^b [Eq. (13)] and lattice relaxations for various Be-H complexes in wurtzite (WZ) and zinc-blende (ZB) GaN. AB stands for antibonding, BC for bond-center sites. For wurtzite, the symbol ∥ denotes configurations with the H-N bond parallel to the c axis ([0001]); the other bond directions are denoted by the symbol ⊥. $\Delta d_{\mathrm{Be(N)}}$ denotes the displacement of the Be(N) atom from its nominal lattice site, expressed as a percentage of the bond length. $d_{\mathrm{N\text{-}H}}$ denotes the N-H distance.

Configuration	E^b (eV)	Δd_{Be}	Δd_{N}	$d_{\mathrm{N\text{-}H}}$ (Å)
WZ, $BC_{\parallel}$	1.81	35.3%	10.8%	1.028
WZ, $BC_{\perp}$	1.68	31.3%	10.6%	1.027
WZ, $AB_{\perp}$	0.95	21.2%	12.8%	1.031
WZ, $AB_{\parallel}$	0.69	23.3%	11.3%	1.030
ZB, BC	1.64	30.4%	13.1%	1.026
ZB, AB	0.90	21.8%	13.9%	1.032

actually succeed in forming strong bonds with N atoms: the Be atom closest to the substitutional site binds with two N atoms, one at 1.54 Å, the other at 1.61 Å, while the Be closer to the interstitial site binds with one N at 1.65 Å and two N [out of the plane in Fig. 6(d)] at 1.61 Å. The Be-Be bond lies in the $(11\bar{2}0)$ plane at an angle of 27° from the c axis.

The formation energy of this complex at $E_F=0$, calculated using Eq. (5), is 0.37 eV for N-rich conditions (see Table I). The binding energy between Be_{int}^{2+} and $\mathrm{Be}_{\mathrm{Ga}}^{-}$ is thus calculated to be 1.35 eV. The formation energy of the complex as a function of E_F is also included in Fig. 3.

We have also investigated the neutral charge state of this complex. No Kohn-Sham states were introduced within the LDA band gap, indicating a shallow donor. We followed the same procedure as for Be_{int} (see Sec. III C 2) to derive the donor ionization energy, resulting in a level about 0.02 eV below the conduction-band minimum.

If Be_{int}-$\mathrm{Be}_{\mathrm{Ga}}$ are incorporated in bulk GaN, one may wonder whether it is possible to subsequently dissociate them. An estimate for the dissociation energy can be obtained by adding the activation energy for interstitial diffusion to the binding energy of the complex. We are allowed to use the lower of the calculated interstitial diffusion barriers, since for

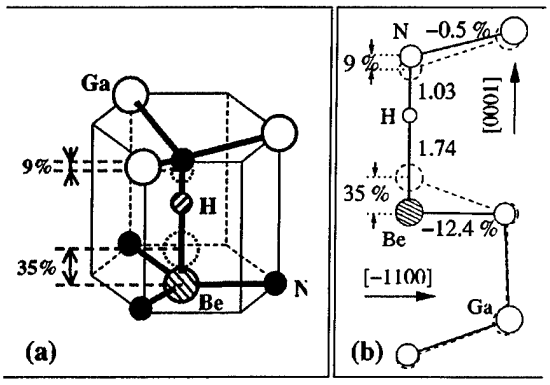


FIG. 9. Schematic illustration of the atomic structure of the Be-H complex. The H atom is located in a bond-center site along the c axis. The $\mathrm{Be}_{\mathrm{Ga}}^{-}$ is relaxed outward by a very large amount, moving it into the plane of its three N neighbors.

TABLE IV. Calculated formation energies for complexes between Be substitutional acceptors ($\mathrm{Be_{Ga}}$) and oxygen substitutional donors ($\mathrm{O_N}$) in wurtzite (WZ) and zinc-blende (ZB) GaN. Nitrogen-rich conditions are assumed, as well as equilibrium with $\mathrm{Be_3N_2}$ and BeO. The formation energies are given for $E_F=0$, i.e., E_F located at the top of the valence band.

	Formation energies (eV)	
	WZ	ZB
$(\mathrm{O_N})^+$	0.24	0.22
$(\mathrm{Be_{Ga}\text{-}O_N})^0$	0.59	0.66
$(\mathrm{Be_{Ga}\text{-}O_N\text{-}Be_{Ga}})^-$	1.97	1.93
$(\mathrm{Be_{Ga}\text{-}O_N\text{-}Be_{Ga}})^0$	1.83	1.78

purposes of breaking the complex the Be interstitial would presumably choose to move in the direction that offers the lowest dissociation barrier. This yields a dissociation energy of 1.35+1.18=2.53 eV. We will discuss in Sec. V what choice of annealing temperatures would result in dissociation of these complexes.

B. Beryllium-hydrogen complexes

Hydrogen is known to passivate dopant impurities in semiconductors, and hydrogen passivation of Mg acceptors in GaN has been well documented.[21–23] In the case of Mg, the (electrically neutral) Mg-H complex was found to have a binding energy of 0.7 eV, with the H atom located in an antibonding (AB) site behind a N neighbor of the acceptor.[22]

We have carried out comprehensive calculations of the Be-H complex, the results of which are summarized in Table III. The binding energy of the complex is defined with respect to an isolated Be acceptor in the negative charge state, and an isolated interstitial H atom in the positive charge state (its lowest-energy state in p-type GaN) far away from the Be acceptor:

$$E_b(\mathrm{Be\text{-}H}) = -E^f[\mathrm{Be\text{-}H}^0] + E^f[\mathrm{Be}_{\mathrm{Ga}}^-] + E^f[\mathrm{H}^+], \quad (13)$$

where the signs have been chosen such that a positive binding energy corresponds to a stable complex.

In the wurtzite structure, the $BC_{\parallel}$ configuration, in which the H atom is located in the bond-center (BC) site between the Be acceptor and its N neighbor along the c axis, is most favorable (highest binding energy). This configuration is illustrated in Fig. 9. The other possible BC configuration, $BC_{\perp}$, in which the Be-H-N bond is pointing along a direction other than [0001], is only 0.1 eV higher in energy. Note that the AB configurations are significantly higher in energy than the BC structures (by more than 0.8 eV). This result is markedly different from the case of Mg-H, where the AB configuration was the more favorable one. Our calculated binding energy for the global minimum of the Be-H complex is in good agreement with that reported by Bernardini *et al.*[6]

We also draw attention to the extremely large atomic relaxations in this complex: the N atom moves outward by about 10% of the Ga-N bond length, and the Be by as much as 35%. Such large relaxations were actually also calculated for the Mg-H complex.

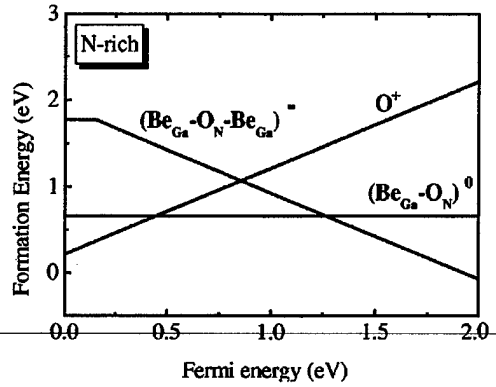


FIG. 10. Calculated defect formation energies as a function of Fermi level for complexes between oxygen and beryllium in zinc-blende GaN, under N-rich conditions.

For all the complexes listed in Table III, the N-H distance is approximately 1.03 Å. Also note that the behavior of the complexes in the zinc-blende phase is qualitatively very similar to that in the wurtzite phase: The BC configuration is significantly more stable than the AB configuration, and the atomic relaxations (particularly of the Be impurity) are very large.

Experimental identification of acceptor-hydrogen complexes often relies on vibrational spectroscopy. We have therefore also calculated the vibrational modes of the Be-H complex, for the most favorable geometry of the complex (i.e., $BC_{\parallel}$). In the harmonic approximation we find a stretch mode at 3460 cm^{-1}, and a wag mode at 570 cm^{-1}. As expected, the stretch mode is representative of a N-H bond, which dominates the bonding in this complex, while the wag modes are low in frequency because the potential energy surface perpendicular to the bond is quite shallow for the BC configuration.[42] More details about the vibrational mode calculation will be published elsewhere.[43]

C. Beryllium-oxygen and Be-O-Be complexes

Complex formation between Be_{Ga} and oxygen has been suggested as an explanation for the high hole conductivities observed in codoping experiments.[18–20] We have carried out calculations for a variety of complexes between oxygen and one or two Be atoms. The results are summarized in Table IV and Fig. 10. In Fig. 10 we focus on zinc blende, which is the phase for which experimental results were reported;[18] however, we notice from Table IV that the results for wurtzite are very similar. The choice of chemical potentials was as follows. We chose N-rich conditions, to allow easy comparison with results presented elsewhere in this paper. For Be, we again assumed its chemical potential to be limited by equilibrium with Be_3N_2. For oxygen, finally, we assumed equilibrium with BeO (for which we used the experimental heat of formation −6.3 eV; Ref. 37). These specific choices are not critical for the qualitative conclusions that we will derive. Note, however, that if we did not include equilibrium with BeO the formation energies of any of the oxygen-containing complexes would be much lower; indeed, the high binding energy of the Be-O bond favors formation of such complexes.

First we examine the $(Be_{Ga}\text{-}O_N)^0$ complex, which is neutral because the O_N^+ donor passivates the Be_{Ga}^- acceptor. The binding energy of the complex is defined with respect to an isolated Be acceptor in the negative charge state, and an isolated O donor in the positive charge state far away from the Be acceptor:

$$E_b(Be_{Ga}\text{-}O_N) = -E^f[Be_{Ga}\text{-}O_N] + E^f[Be_{Ga}^-] + E^f[O_N^+], \tag{14}$$

where the signs have been chosen such that a positive binding energy corresponds to a stable complex. From the calculated formation energies, we obtain a binding energy for $(Be_{Ga}\text{-}O_N)$ of 1.33 eV in WZ and 1.23 eV in ZB. This is quite a large value, and indicates that these complexes are likely to form.

Interestingly, in this Be-O complex the Be and O atoms move away from their substitutional lattice sites, in opposite directions. Be moves by a distance equal to 22% of the Ga-N bond length, and oxygen moves by 12% of the bond length. The final distance between the Be and O atoms is such that only a very weak bond is present between the impurities. We propose the following tentative explanation for the observed tendency for Be and O to move away from one another. The natural Be-O bond length is very short; for instance, the bond distance in BeO distance is 1.65 Å.[44] In order to maintain this preferred Be-O distance for Be and O nearest neighbors in GaN, a large displacement of Be away from its N neighbors (and/or of O away from its Ga neighbors) would be required, in turn leading to significant relaxations of those N and Ga atoms. Such large relaxations are energetically costly. Our result indicates that it is actually preferable for Be and O to both optimize their bond lengths to their surrounding N and Ga neighbors, and essentially "give up" on forming a Be-O bond. Indeed, both Be and O seem to be comfortable when bonded to only three neighbors in GaN; for Be, this tendency has been discussed above in Sec. III A; for O it has been observed in the context of *DX*-center formation.[45]

Turning now to $(Be_{Ga}\text{-}O_N\text{-}Be_{Ga})$ complexes, we note from Table IV and Fig. 10 that these have formation energies that are slightly higher than those of isolated Be_{Ga} acceptors. We therefore do not expect a higher solubility of Be based upon formation of such complexes. Our calculated ionization energy of the Be-O-Be complex is 0.14 eV in WZ and 0.15 eV in ZB; within our error bars, this is comparable to that of the isolated substitutional Be. As to the structure of the $(Be_{Ga}\text{-}O_N\text{-}Be_{Ga})$ complex, we again find that one Be-O bond is effectively broken. In the WZ structure, the configuration with no Be-O bonds along the *c* axis was found to be favored. It is informative to calculate the binding energy for the process whereby an additional Be_{Ga}^- is attached to a $(Be_{Ga}\text{-}O_N)$ complex, resulting in a $(Be_{Ga}\text{-}O_N\text{-}Be_{Ga})^-$ complex. Using the tabulated formation energies we find a binding energy of 0.30 eV in WZ and 0.40 eV in ZB. This binding energy is significantly smaller than the binding energy found above for the Be-O complex. This indicates that once a Be-O complex is formed the driving force for attaching a second Be atom is quite small.

V. DISCUSSION

A. Incorporation of Be in GaN

Our calculated formation energy for Be_{Ga} in GaN is slightly lower than that of Mg_{Ga}, which is the most widely used acceptor.[5,35] Comparing our results for formation energies (which determine solubilities) with previous calculations by Bernardini *et al.*,[6] we find reasonable agreement for the formation energy of the substitutional acceptor (allowing for differences relating to the choice of atomic chemical potentials, etc.). The formation energies obtained in Ref. 6 for *interstitial* Be are significantly larger than those found in the present work (by more than 2.0 eV). We are unable to provide an explanation for this discrepancy.

In Sec. III A we pointed out that by moving from N-rich to Ga-rich conditions the formation energy of Be_{Ga} increases; indeed, we chose to list results for N-rich conditions precisely because that leads to the highest possible solubility for Be_{Ga}. It turns out that moving toward Ga-rich conditions also *decreases* the formation energy of Be_{int} (because we are assuming equilibrium with Be_3N_2). Optimizing Be incorporation (and suppressing Be interstitial formation) would thus require N-rich conditions. Unfortunately, growing high-quality GaN usually requires Ga-rich growth conditions,[46] which are less favorable for Be incorporation.

The large binding energies found for Be-H complexes (see Table III) are in agreement with the results of Ref. 6. If H is incorporated during growth, Be-H complexes are sure to form. Incorporation of H may happen unintentionally, such as during MOCVD (metal-organic chemical vapor deposition) growth or when NH_3 is used as a source gas in MBE. Intentional exposure to hydrogen may also be obtained by introducing atomic hydrogen in the chamber during growth. The large binding energy of the Be-H complex results in a low formation energy, which would promote incorporation of Be and effectively increase the solubility. We previously discussed this mechanism in the case of Mg doping.[23]

As in the case of Mg doping, the resulting acceptor-hydrogen complexes need to be broken and the hydrogen removed from the vicinity of the acceptors before electrical activation of the *p*-type layer can be achieved. The binding energy of 1.81 eV may seem to make it difficult to break Be-H complexes. An estimate for the dissociation energy of these complexes can be obtained by adding the activation energy for H^+ diffusion, 0.7 eV,[22] to the binding energy of the Be-H complex. This yields a dissociation barrier of about 2.51 eV. Although this is a high value, it should not be a problem to dissociate the complexes at the temperatures typically used for activation of the *p*-type layers. Indeed, it is well known that hydrogen-acceptor complexes in Si, which have a dissociation energy of about 1.2 eV, can be broken up at temperatures of about 150 °C. Assuming similar dissocation kinetics, complexes with a barrier twice as high should therefore dissociate at a temperature of $2\times(150+273)$ K, or 573 °C. This conclusion is in agreement with the assessment of Bernardini *et al.*[6]

Experiments on Be incorporation in GaN are still scarce. Incorporation of Be in GaN has so far been performed during MBE growth[7–15] and using ion implantation.[16,17] The earliest report of Be doping was by Ilegems and Dingle,[47] who incorporated Be during vapor growth of GaN. The introduction of Be resulted in a deep yellow-green luminescence band centered around 2.2, which is most likely defect related. Sánchez *et al.*[7–9] incorporated Be in WZ GaN grown by MBE on Si substrates. SIMS (secondary ion mass spectroscopy) showed the presence of Be, but due to a lack of calibration no concentrations could be inferred. Conductivity measurements of the Be-doped layer were impossible, due to the electrical activity of Ga diffused into the underlying Si substrate. Finally, Cheng *et al.*[10,11] reported Be incorporation in ZB and WZ GaN grown by MBE on GaAs substrates. SIMS indicated Be incorporation, but again no calibration was available. For high Be concentrations, the SIMS profiles showed evidence of Be diffusion through an undoped GaN buffer layer and into the underlying GaAs substrate.

Ion implantation was used by Ronning *et al.*[16,17] to incorporate Be in MOCVD-grown GaN, followed by annealing at 1100 °C for at least one hour under an atomic nitrogen flux. Beryllium concentrations of more than 10^{17} cm^{-3} were obtained as measured by SIMS. No long-range diffusion of Be was observed, but this result could merely reflect the presence of implantation damage.

B. Ionization energies

Our calculated ionization energy for the Be acceptor in WZ GaN is 170 meV. This value is slightly lower (by about 30 meV) than our previously calculated value for Mg in GaN.[23] We emphasize that the error bar on these values (which we estimate to be at least 100 meV) does not allow us to draw firm conclusions about the magnitude of the ionization energy. Still, the similarity of the values for Be and Mg is in line with the expectation that the ionization energy for these shallow acceptors is largely determined by intrinsic properties of the semiconductor, such as effective masses and dielectric constants. Indeed, predictions from effective-mass theory[48] for ionization energies of substitutional acceptors in WZ GaN produce a value for Be between 185 and 233 meV—only slightly lower than the calculated value for Mg.

Our result for the ionization energy is larger than the 60 meV reported by Bernardini *et al.*[6] for Be in WZ GaN. Their value for Be in ZB GaN was even lower, the acceptor level being resonant with the valence band. One potential explanation for the discrepancy is the smaller unit cell (32 atoms) used in their calculations—although our own results for 32-atom cells produce ionization energies quite close to the converged values. It is also possible that the correction term E_{corr}, discussed in Sec. II D, was not taken into account. This correction lowers the formation energy of the neutral charge state, and hence increases the ionization energy. Neglect of the correction would thus result in artificially low acceptor ionization energies.

In Sec. III A we reported that the large lattice relaxation for Be_{Ga}^0 would result in a Franck-Condon shift of about 100 meV, this being the energy difference between Be_{Ga}^- with a configuration fixed to be equal to that of the Be_{Ga}^0 neutral acceptor, and Be_{Ga}^- in its lowest-energy, relaxed configura-

tion (see Fig. 2). We thus expect the optical ionization energy to be higher than the thermal ionization energy by 100 meV (see Sec. II F). Thermal ionization energies for Be acceptors in GaN have not yet been reported, due to a lack of electrical conductivity data for Be-doped samples. Photoluminescence spectroscopy has been performed by several groups, however, and various estimates for the optical ionization energy have been published.

Sánchez *et al.*[7,8] observed a PL line at 3.384 eV, which they attributed to a donor-acceptor transition. This led to an estimate for the optical ionization energy of 90 meV. This result agrees with the work of Dewsnip *et al.*,[13] who reported a PL line at 3.376 eV; again assuming this to be a donor-acceptor transition, they calculated an ionization energy of 90–100 meV.

Other groups, however, have reached different conclusions regarding the optical ionization energy. Cheng *et al.*[11] observed a PL line at 2.90 eV in their zinc-blende material, and estimated an acceptor ionization energy of 200 meV. Salvador *et al.*[14] observed a PL peak at 390 nm (3.187 eV), and inferred from this an acceptor ionization energy of 250 meV. They also reported that hot-probe measurements indicated the presence of *p*-type conductivity. Ronning *et al.*[16] observed a PL line at 3.35 eV in their Be-implanted material. They attributed the line to band-to-acceptor (eA) transitions, and extracted an ionization energy of 150 ± 10 meV.

The reported values for the optical ionization energy thus cover a wide range, from 90 meV up to 250 meV. Based on our calculated Franck-Condon shift of 100 meV, we think it is very unlikely that the optical ionization energy of the substitutional acceptor could be as low as 90 meV, or even 150 meV. Indeed, this would result in a thermal ionization energy smaller than 50 meV, which conflicts with theoretical predictions and for which there is currently no experimental evidence. We suggest that the photoluminescence data producing optical ionization energies below 150 meV be reexamined, in terms of their attribution to an isolated Be substitutional acceptor and/or the interpretation of the transition responsible for the emission. One could speculate that the observed transition may involve Be interstitials, for which we calculated a $++/+$ ionization level about 90 meV below the conduction band (Sec. III C 2).

C. Codoping with oxygen

Brandt *et al.*[18,19] reported high levels of Be incorporation in cubic GaN grown by MBE on semi-insulating GaN substrates. Oxygen was introduced as a codopant. SIMS indicated concentrations around 5×10^{20} cm^{-3} for both Be and O. No significant diffusion of Be into the GaN buffer layer was observed; formation of Be-O complexes was mentioned as a potential reason for the lack of diffusion. Room-temperature conductivities as high as $50/\Omega$ cm were reported. Hall measurements (performed after removal of the substrate) produced hole concentrations of up to 5×10^{18} cm^{-3}, and mobilities of up to 150 cm^2/V s.

Yamamoto and Katayama-Yoshida proposed an explanation for these codoping results in terms of Be_{Ga}-O_N-Be_{Ga} complexes.[20] They performed first-principles calculations, but their arguments were mainly based on trends in the electrostatic (Madelung) energy.

We reported our full energy-minimization calculations for the Be-O-Be complexes in Sec. IV C. We found that the formation energy of these complexes is not lower than that of isolated Be acceptors (at least if equilibrium with the proper solubility-limiting phases is taken into account). Formation of Be-O complexes (which are electrically neutral) is energetically quite favorable, but attaching a second Be atom to such a complex is only marginally favored. Finally, even if Be-O-Be complexes form, their ionization energy does not seem to be noticeably lower than that of isolated Be acceptors (within the error bars of our calculations). We therefore doubt that formation of Be-O-Be complexes is a viable approach to increasing *p*-type doping of GaN, or provides an explanation for the experimentally observed *p*-type conductivity in codoped samples.

D. Compensation and how to avoid it

Our calculated formation energies for Be interstitials clearly indicate that they would easily incorporate as compensating centers. In WZ GaN, our calculated formation energy of Be_{int}^{2+} is $0.04\ \text{eV}+2E_F$. To put this into perspective, this formation energy is lower than that of the nitrogen vacancy (which is the dominating intrinsic defect in *p*-type GaN), at least for Fermi-level positions within 1 eV of the top of the valence band.

We note that codoping with hydrogen during growth offers a very effective means of avoiding compensation. Indeed, hydrogen itself is a donor with a very low formation energy,[22,23] and hence is more likely to incorporate as a compensating center than the Be interstitial—especially when the large binding energy (and hence low formation energy) of the Be-H complex is taken into account. Compensation by Be interstitials would thus be effectively suppressed in the presence of hydrogen. The hydrogen would need to be removed in a postgrowth annealing process before activation of the acceptors could be achieved, but as discussed in Sec V A that should not pose any problems. We emphasize that the success of codoping crucially depends on the ability to remove the codopant after the layers have been grown. This is possible in the case of hydrogen, but not in the case of other elements (such as oxygen) that have been suggested as codopants. This issue has been discussed in Ref. 23.

If, for any reason, Be interstitials are not incorporated *during* growth, they might still diffuse into the bulk *after* growth. Our calculated diffusion barrier for motion along the *c* axis is 2.90 eV; we estimated in Sec. III C 5 that this would result in significant diffusion at temperatures above 850 °C.

Any Be interstitials that are incorporated are likely to bind to Be_{Ga} acceptors; indeed, their diffusion barrier (at least perpendicular to *c*) is quite low, and Be_{int}^{2+} is electrostatically attracted to Be_{Ga}^{-}. We found the binding energy of such complexes to be 1.35 eV, and in Sec. IV A we estimated the dissociation energy to be 2.53 eV. This is very similar to the dissociation energy of Be-H complexes, and thus by the same arguments as used in Sec. V A annealing temperatures

of 600 °C or higher should be adequate to break such complexes. The resulting isolated Be interstitials still need to be "neutralized" in order to achieve activation of the Be acceptors. This can be achieved either by completely removing the Be interstitials from the *p*-type layer, or by neutralizing the interstitial donors, for example at extended defects. If lateral diffusion suffices to achieve this neutralization, then the same temperatures as required for breaking of the complexes should suffice to activate the *p*-type layer. However, if diffusion parallel to *c* is necessary (for instance, to remove the Be interstitials through the surface of the wurtzite layer), then higher temperatures may be required due to the higher diffusion barrier for interstitials in that direction.

VI. SUMMARY AND OUTLOOK

We have reported comprehensive first-principles results for Be in GaN. The formation energy of Be acceptors is slightly lower than that of Mg, indicating that Be would exhibit a higher solubility. We also predict a slightly lower thermal ionization energy for Be. The optical ionization energy is calculated to be higher than the thermal ionization energy by 100 meV, due to a large lattice relaxation of the Be acceptor in its neutral charge state.

The calculations show that Be interstitials have a low formation energy, and are hence likely to incorporate during growth. Be interstitials are double donors and hence cause compensation. We mapped out the total-energy surface for diffusion of Be interstitials, finding a highly anisotropic behavior, with diffusion along the *c* axis having a much larger barrier (2.9 eV) than diffusion in planes perpendicular to the *c* axis (1.2 eV). The interstitials may bind to substitutional acceptors, with a binding energy of 1.35 eV and a dissociation energy of 2.53 eV.

In the presence of hydrogen, beryllium-hydrogen complexes may form with a binding energy of 1.81 eV and a dissociation energy of 2.51 eV. Formation of such complexes would result in a higher solubility of substitutional Be acceptors, and suppression of compensation by Be interstitial donors. A postgrowth activation anneal is then required to remove the hydrogen from the vicinity of the Be acceptors.

Based on our results and analysis, we can suggest several experiments that could shed further light on the behavior of Be in GaN. Obviously, an experimental determination of the thermal ionization energy would be very valuable. Further investigations of the optical ionization energy, particularly in light of the predicted large Franck-Condon shift, are also desired.

To study the diffusion of interstitial Be, diffusion of Be into GaN from a Be source on the surface could be attempted. For WZ GaN, this would probably mainly involve diffusion along the *c* axis, which we predicted to have a high barrier. If diffusion along extended defects could be excluded, we would therefore expect only very slow Be diffusion for this experimental configuration. It would be highly interesting if an experimental geometry could be conceived in which interstitial diffusion perpendicular to the *c* axis could be studied, which we predict to have a relatively low barrier.

If Be interstitials are incorporated during growth, then postgrowth annealing experiments would be very informative. Indeed, we estimated that typical annealing temperatures of 600 °C or higher would result in dissociation of Be_{int}-Be_{Ga} complexes and rapid in-plane diffusion of Be interstitials. Diffusion parallel to *c* may require higher temperatures.

Finally, the role of hydrogen is likely to be a very beneficial one: its intentional or unintentional incorporation should be monitored, formation of Be-H complexes may be detected (for instance by using vibrational spectroscopy and comparing with our predicted vibrational modes), and dissociation of such complexes by thermal annealing should be attempted.

In conclusion, our first-principles studies have elucidated the attractive aspects of beryllium as an acceptor in GaN, but also the potential problems, which are mainly related to the incorporation of Be interstitials. We hope that our comprehensive investigation of the properties of the Be dopant will prove useful in guiding and interpreting experiments.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research, Contract No. N00014-99-C-0161 and by the Air Force Office of Scientific Research, Contract No. F4920-00-C-0019. We thank J. E. Northrup, L. Romano, N. Johnson, M. Kneissl, P. Kiesel, and T. H. Myers for useful discussions. C.V.d.W. is grateful to the Fritz-Haber-Institut and Paul-Drude-Institut, Berlin, for their hospitality, and to the Alexander von Humboldt Foundation for financial support.

*Email address: vandewalle@parc.xerox.com

[1] W. Götz, R. S. Kern, C. H. Chen, H. Liu, D. A. Steigerwald, and R. M. Fletcher, Mater. Sci. Eng., B **59**, 211 (1999).

[2] D. P. Bour, H. F. Chung, W. Gtz, L. Romano, B. S. Krusor, D. Hofstetter, S. Rudaz, C. P. Kuo, F. A. Ponce, N. M. Johnson, M. G. Craford, and R. D. Bringans, in *III-V Nitrides*, edited by F. A. Ponce, T. D. Moustakas, I. Akasaki, and B. Monemar, Mater. Res. Soc. Symp. Proc. No. **449** (Materials Research Society, Pittsburgh, 1997), p. 509.

[3] J. Neugebauer and C. G. Van de Walle, in *Gallium Nitride and Related Materials*, edited by R. D. Dupuis, J. A. Edmond, F. A. Ponce, and S. Nakamura, Mater. Res. Soc. Symp. Proc. No. **395** (Materials Research Society, Pittsburgh, 1996), p. 645.

[4] V. Fiorentini, F. Bernardini, A. Bosin, and D. Vanderbilt, in *Proceedings of the 23rd International Conference on the Physics of Semiconductors*, edited by M. Scheffler and R. Zimmermann (World Scientific, Singapore, 1996), p. 2877.

[5] J. Neugebauer and C. G. Van de Walle, J. Appl. Phys. **85**, 3003 (1999).

[6] F. Bernardini, V. Fiorentini, and A. Bosin, Appl. Phys. Lett. **70**, 2990 (1997).

[7] F. J. Sánchez, F. Calle, M. A. Sánchez-García, E. Calleja, E. Muñoz, C. H. Molloy, D. J. Somerford, J. J. Serrano, and J. M. Blanco, Semicond. Sci. Technol. **13**, 1130 (1998).

[8] R. J. Sánchez, F. Calle, M. A. Sánchez-García, E. Calleja, E. Muñoz, C. H. Molloy, D. J. Somerford, F. K. Koschnick, K. Michael, and J.-M. Spaeth, MRS Internet J. Nitride Semicond. Res. **3**, 19 (1998).

[9] M. A. Sánchez-García, E. Calleja, F. J. Sánchez, F. Calle, E. Monroy, D. Basak, E. Muñoz, C. Villar, A. Sanz-Hervas, M. Aguilar, J. J. Serrano, and J. M. Blanco, J. Electron. Mater. **27**, 276 (1998).

[10] T. S. Cheng, C. T. Foxon, L. C. Jenkins, S. E. Hooper, D. E. Lacklison, J. W. Orton, B. Ya Ber, A. V. Merkulov, and S. V. Novikov, Semicond. Sci. Technol. **11**, 538 (1996).

[11] T. S. Cheng, S. E. Hooper, L. C. Jenkins, C. T. Foxon, D. E. Lacklison, J. D. Dewsnip, and J. W. Orton, J. Cryst. Growth **166**, 597 (1996).

[12] D. J. Dewsnip, A. V. Andrianov, I. Harrison, D. E. Lacklison, J. W. Orton, J. Morgan, G. B. Ren, T. S. Cheng, S. E. Hooper, and C. T. Foxon, Semicond. Sci. Technol. **12**, 55 (1997).

[13] D. J. Dewsnip, A. V. Andrianov, I. Harrison, J. W. Orton, D. E. Lacklison, G. B. Ren, S. E. Hooper, T. S. Cheng, and C. T. Foxon, Semicond. Sci. Technol. **13**, 500 (1998).

[14] A. Salvador, W. Kim, Ö. Aktas, A. Botchkarev, Z. Fan, and H. Morkoç, Appl. Phys. Lett. **69**, 2692 (1996).

[15] T. H. Myers *et al.* (unpublished).

[16] C. Ronning, E. P. Carlson, D. B. Thomson, and R. F. Davis, Appl. Phys. Lett. **73**, 1622 (1998).

[17] C. Ronning, K. J. Linthicum, E. P. Carlson, P. J. Hartlieb, D. B. Thomson, T. Gehrke, and R. F. Davis, MRS Internet J. Nitride Semicond. Res. **4S1**, G3.17 (1999).

[18] O. Brandt, H. Yang, H. Kostial, and K. H. Ploog, Appl. Phys. Lett. **69**, 2707 (1996).

[19] K. H. Ploog and O. Brandt, J. Vac. Sci. Technol. A **16**, 1609 (1998).

[20] T. Yamamoto and H. Katayama-Yoshida, in *Proceedings of the 19th International Conference on Defects in Semiconductors, Aveiro, Portugal, 1997*, edited by G. Davies and M. H. Nazaré [Mater. Sci. Forum **258-263**, 1185 (1997)].

[21] S. Nakamura, N. Iwasa, M. Senoh, and T. Mukai, Jpn. J. Appl. Phys., Part 1 **31**, 1258 (1992).

[22] J. Neugebauer and C. G. Van de Walle, Phys. Rev. Lett. **75**, 4452 (1995).

[23] J. Neugebauer and C. G. Van de Walle, Appl. Phys. Lett. **68**, 1829 (1996).

[24] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964); W. Kohn and L. J. Sham, *ibid.* **140**, A1133 (1965).

[25] M. Bockstedte, A. Kley, J. Neugebauer, and M. Scheffler, Comput. Phys. Commun. **107**, 187 (1997).

[26] J. Neugebauer and Chris G. Van de Walle, in *Diamond, SiC and Nitride Wide Bandgap Semiconductors*, edited by C. H. Carter, Jr., G. Gildenblat, S. Nakamura, and R. J. Nemanich, Mater. Res. Soc. Symp. Proc. No. **339** (Materials Research Society, Pittsburgh, 1994), p. 687.

[27] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).

[28] C. Stampfl and C. G. Van de Walle, Phys. Rev. B **59**, 5521 (1999).

[29] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[30] This correction was also taken into account in a previous publication (Ref. 5); however, there the correction was applied to the *negative* charge state, rather than the *neutral* charge state. This accounts for some of the differences between the results presented in Ref. 5 and in the present paper.

[31] D. B. Laks, C. G. Van de Walle, G. F. Neumark, P. E. Blöchl, and S. T. Pantelides, Phys. Rev. B **45**, 10 965 (1992).

[32] *CRC Handbook of Chemistry and Physics*, 73rd ed., edited by David R. Lide (CRC Press, Boca Raton, FL, 1992), pp. 5–18.

[33] J. E. Northrup (private communication).

[34] *Lange's Handbook of Chemistry*, 14th ed., edited by J. A. Dean (McGraw-Hill, New York, 1992), p. 6.72.

[35] Note that if we use the *experimental* value of ΔH_f [Be_3N_2] to determine μ_{Be}, the calculated formation energy will be 0.2 eV larger than if we use the *calculated* value of ΔH_f [Be_3N_2]. This small difference does not alter any of our conclusions, but may be kept in mind when comparing with the formation energies of other acceptors as calculated in Ref. 5, where experimental values for ΔH_f of the solubility-limiting compounds were used.

[36] *Periodic Table of the Elements* (Sargent-Welch Scientific Company, Skokie, 1980).

[37] *Lange's Handbook of Chemistry* (Ref. 34), p. 4.13.

[38] J. E. Northrup (unpublished).

[39] R. W. G. Wyckoff, *Crystal Structures*, 2nd ed. (Interscience Publishers, New York, 1964), Vol. 2, pp. 32–33.

[40] D. J. Chadi and K. J. Chang, Phys. Rev. Lett. **61**, 873 (1988).

[41] J. Neugebauer and C. G. Van de Walle, Appl. Phys. Lett. **69**, 503 (1996).

[42] C. G. Van de Walle, P. J. H. Denteneer, Y. Bar-Yam, and S. T. Pantelides, Phys. Rev. B **39**, 10 791 (1989).

[43] S. Limpijumnong and C. G. Van de Walle (unpublished).

[44] R. W. G. Wyckoff, *Crystal Structures* (Ref. 39), Vol. 1, pp. 108–110.

[45] C. G. Van de Walle, Phys. Rev. B **57**, R2033 (1998).

[46] E. J. Tarsa, B. Heying, X. H. Wu, P. Fini, S. P. DenBaars, and J. S. Speck, J. Appl. Phys. **82**, 5472 (1997).

[47] M. Ilegems and R. Dingle, J. Appl. Phys. **44**, 4234 (1973).

[48] F. Mireles and S. E. Ulloa, Phys. Rev. B **58**, 3879 (1998).

Theoretical investigation of native defects, impurities, and complexes in aluminum nitride

C. Stampfl[1] and C. G. Van de Walle[2]
[1]*Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208-3112*
[2]*Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, California 94304*
(Received 1 November 2001; published 15 April 2002)

We have performed density-functional pseudopotential calculations to investigate the electronic structure, atomic configurations, and formation energies of native point defects and impurities in AlN. For the native defects, the nitrogen vacancy has the lowest formation energy in *p*-type material and the aluminum vacancy has the lowest formation energy in *n*-type material. Under *n*-type conditions the formation energy of the nitrogen vacancy is high, indicating that it will not occur in high concentrations. We find that the nitrogen vacancy exhibits a *different* behavior in the zinc-blende and wurtzite structures with respect to the higher-lying defect-induced level: in zinc-blende materials, this level is a resonance in the conduction band causing the vacancy to act as a shallow donor, while in wurtzite the level lies well below the conduction-band edge causing the vacancy to act as a deep donor. In the zinc-blende structure we find, in addition, that the aluminum interstitial has a low formation energy in *p*-type material. The results indicate that these defects could be important compensation centers; we discuss this in relation to the dopant impurities O, Si, and Mg. We also investigate MgO and Mg_2O_2 impurity complexes. A comparison between results obtained using the local-density approximation and the generalized-gradient approximation for the exchange-correlation functional shows that the results are qualitatively very similar.

DOI: 10.1103/PhysRevB.65.155212 PACS number(s): 61.72.Ji, 71.15.Nc, 71.55.Eq

I. INTRODUCTION

Group-III nitride materials are being intensively researched due to their demonstrated success in optoelectronic devices and their further potential in this area.[1–3] For example, highly efficient blue and green light-emitting diodes have been fabricated,[4] and nitride-based laser diodes have been developed.[5,6] Considerable effort is being focused on obtaining an understanding of the microscopic electronic and atomic structure which will afford further improvement and control of device performance. An important step in this direction is the study of native point and impurity defects. While there have been several theoretical studies of point defects in GaN, less work has been done on AlN. It is, however, important to understand the behavior of defects in AlN in addition to GaN since, e.g., the cladding and confinement layers which surround the active region of a laser diode are typically $Al_xGa_{1-x}N$, for which it must be possible to perform both *n*- and *p*-type doping. Successful *p*- and *n*-type doping of AlN could also lead to important applications for this material. In GaN, one factor limiting *p*-type conductivity is self-compensation, apparently resulting from the formation of donor-acceptor pairs;[7,8] a similar problem could also occur for AlN, which to date has resisted attempts at doping.

With respect to theoretical studies of native defects and impurities in AlN, we mention the following: Jenkins and Dow[9] considered defects in the neutral charge state using an empirical tight-binding method. Defect levels were reported, but no calculation of formation energies was attempted. More recently, first-principles pseudopotential calculations were performed by Mattila and Nieminen,[10] who reported formation energies of various charge states of the N and Al vacancies and the O impurity in zinc-blende AlN, as well as the N antisite.[11] It was reported that the nitrogen vacancy (V_N) induced two states in the band gap, one doubly occupied state lying above the valence-band maximum (VBM) and a higher-lying level that could be occupied with one or three electrons (the neutral and singly negative charge states being unstable). The nature of the latter state is different from what was reported in our subsequent work (Refs. 12 and 13), where we found that the upper V_N-induced state was a resonance in the conduction band. As we explain below, it appears that the calculations of Ref. 10 erroneously occupied the lowest conduction band rather than the defect state. Recently, Fara *et al.*[14] studied Al and N vacancies as well as selected impurities in *wurtzite* AlN for various charge states using the pseudopotential method. With respect to the nitrogen vacancy, they found the upper defect state to be deep in the band gap, i.e., about 1–3 eV below the conduction-band minimum. Our systematic calculations will indeed confirm this difference between zinc-blende and wurtzite material, the nitrogen vacancy acting as a shallow donor in the former, but as a deep donor in the latter.

Regarding impurities, C, Si, and Ge (Ref. 15) and Si and Ge (Ref. 16) in wurtzite GaN and AlN were investigated in the neutral charge state by Boguslawski *et al.* and by Park and Chadi, respectively. Finally, Gorczyca *et al.*[17] investigated native point and impurity defects using the Green's-function technique based on the linear muffin-tin orbital method in the atomic sphere approximation with no atomic relaxations included; here the emphasis was on the influence of hydrostatic pressure on the defect levels with no formation energies being reported. In a later publication these authors studied native defects and the C, Zn, and Mg impurities in cubic (zinc-blende) AlN and GaN in the neutral charge state, this time taking atomic relaxation into account and reporting formation energies.[18] With respect to the study of Mg and O defect complexes, Gorczyca *et al.*[19] investigated Mg-O (and

 ©2002 The American Physical Society

Mg-V_N) structures in (cubic) GaN, and proposed that such complexes could play a role in self-compensation when attempting p-type doping, which could also have implications for AlN. In earlier work, Van de Walle *et al.*[20] reported a binding energy of 0.6 eV for the Mg-O complex in both (zinc-blende) GaN and AlN.

In the present paper we perform a systematic investigation of the electronic and atomic structure, as well as defect formation energies, of native point defects, selected impurities (Si, O, and Mg), and defect complexes MgO and Mg_2O_2 in AlN using first-principles pseudopotential calculations. Some of these results were briefly reported earlier.[12,13] The paper is organized as follows: In Sec. II we describe the calculation method and in Sec. III results for the native defects are presented. Section IV describes results for the impurities and defect complexes, and Sec. V contains the conclusion.

II. CALCULATION METHOD

The calculations are performed using density-functional theory (DFT) and the pseudopotential plane wave method.[21] We use the supercell approach, and employ a tight-binding initialization scheme for the electronic wave functions to improve convergence. Both the local-density approximation[22] (LDA) and the generalized-gradient approximation[23] (GGA) for the exchange-correlation functional are used. In the literature the GGA is considered as a possible alternative for an improved treatment of the exchange-correlation functional over the LDA, and was found to yield a better description of a number of physical properties.[24] To date, there have been very few defect calculations that employ the GGA. AlN can assume both the zinc-blende and wurtzite structures, the latter being the stable phase and the former being stabilized through epitaxial growth.[25] The majority of our calculations are performed using 32-atom zinc-blende supercells and an energy cutoff of 40 Ry with two special **k** points in the irreducible part of the Brillouin zone.[26] We also performed some calculations for nitrogen and aluminum vacancies, which are the dominant native defects, using 32- and 72-atom wurtzite supercells with two and three special **k** points in the Brillouin zone,[27] respectively; the results obtained with the larger cell were very similar to the 32-atom cell. The positions of the atoms around the defects are relaxed using a damped dynamics approach.[21] The pseudopotentials are created using the scheme of Troullier and Martins,[28] and for the GGA calculations we include the GGA in the creation of the pseudopotential as well as in the self-consistent total-energy calculations, so that the description is consistent.[29,30] In earlier work[31] we thoroughly tested the performance of these pseudopotentials, and found bulk properties in very good agreement with other available *ab initio* results. These potentials have also been used in our other recent studies of group-III-nitride systems.[12,13,32,33] While the GGA improved the properties for N_2, bulk Al, and AlN, the improvement for GaN was less significant and for InN, we found that the GGA exhibited an underbinding.

The formation energy of a native point defect in charge state q in AlN is expressed as

$$E^f(q)=E^{\rm tot}_{\rm defect}(q)-E^{\rm tot}({\rm bulk})-n_{\rm Al}\mu_{\rm Al}-n_{\rm N}\mu_{\rm N}+qE_F, \quad (1)$$

where $E^{\rm tot}_{\rm defect}(q)$ is the total energy of the supercell containing the defect, $E^{\rm tot}$(bulk) is the total energy of a similar supercell containing the pure bulk crystal, $n_{\rm Al}$ and $n_{\rm N}$ are the number of Al and N atoms, respectively, that are added to or taken from the bulk crystal in order to create the defect, and $\mu_{\rm Al}$ and $\mu_{\rm N}$ are the corresponding chemical potentials. E_F is the Fermi energy, which is set to zero at the valence-band maximum. We note that for charged defects, a proper lineup of the band structure of the defect supercell with that of the bulk is required. This is done using the difference in the average potential around an atom in a bulklike environment as a measure of the shift needed to line up the band structure (see Refs. 34 and 35 for details).

A low value of the formation energy indicates a high equilibrium concentration of defects, while a high value implies a low concentration. That is, for a given temperature and in the limit of low concentrations, one can assume that the defect concentrations obey an Arrhenius type of behavior, which means that the logarithms of the concentrations are proportional to the formation energies.

The chemical potentials depend on the experimental conditions under which the material is grown. In order to determine these quantities, we invoke the relationship $\mu_{\rm Al}+\mu_{\rm N}=\mu_{\rm AlN}$, assuming both species are in thermal equilibrium with AlN. Furthermore, the chemical potentials must satisfy the boundary conditions: $\mu_{\rm N}<1/2\mu_{\rm N_2}$ and $\mu_{\rm Al}<\mu_{\rm Al(bulk)}$ (if this were not the case, then AlN would be thermodynamically unstable with respect to the formation of N_2 molecules or bulk Al). When imposing certain growth conditions [nitrogen-rich ($\mu_{\rm N}=1/2\mu_{\rm N_2}$) or aluminum-rich ($\mu_{\rm Al}=\mu_{\rm Al(bulk)}$) conditions], the chemical potential for the other species can be determined from the above (thermal equilibrium) relationship.

For the case of an impurity of species X, a term $-n_X\mu_X$ will appear, in addition, on the right-hand side of Eq. (1). In the present work we consider O, Si, and Mg point impurities, and MgO and Mg_2O_2 defect complexes. The atomic chemical potentials are assumed to be determined by equilibrium with Al_2O_3, Si_3N_4, and Mg_3N_2 (leading to maximum solubility). For Al_2O_3 (sapphire) we use the calculated heat of formation of 18.4 eV (LDA) and 16.5 eV (GGA). To be consistent with the defect calculations, these values are obtained at the same energy cutoff, and we optimized the lattice constant. The experimental value is 17.24 eV.[36] For Si_3N_4 and Mg_3N_2 we use the experimental values of the heat of formation (4.81 and 7.70 eV, respectively[36]) together with the energy of the calculated elemental phase.

In the zinc-blende structure, AlN has an *indirect* band gap; we calculate the minimum gap (indirect $\Gamma_{15}-X_1$) to be 3.15 eV. The corresponding experimental value is not well established, but it will be notably larger than the theoretical one

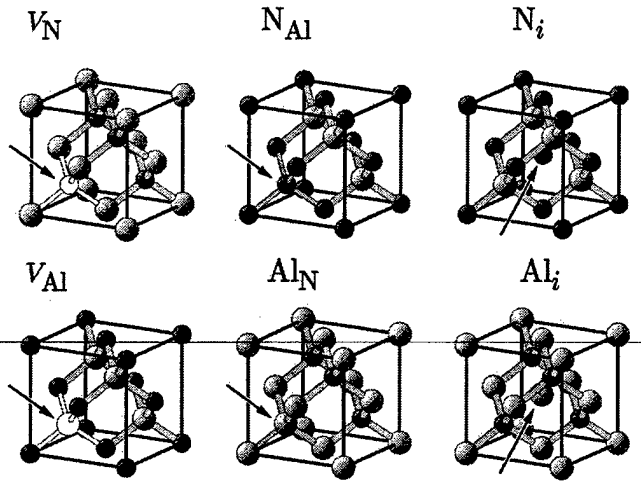


FIG. 1. Atomic geometry of native point defects in zinc-blende AlN in the ideal (unrelaxed) structure. Large gray and small dark spheres represent Al an N atoms, respectively. Vacancies are indicated by the white circles.

reported here due to the well-known tendency for DFT-LDA (and DFT-GGA) calculations (and even exact DFT (Ref. 37)] to underestimate band gaps. In the present work we consider E_F, spanning the indirect band gap of ≈ 5.0 eV as obtained from GW calculations.[38] The calculated wurtzite (direct) band gap is 4.74 eV, and the experimental value is 6.28 eV (see, e.g., Ref. 31).

III. NATIVE POINT DEFECTS

The ideal (unrelaxed) atomic geometry of the point defects in zinc-blende AlN is shown in Fig. 1: N and Al vacancies (V_N and V_{Al}), N and Al antisites (N_{Al} and Al_N), and N and Al interstitials (N_i and Al_i). The associated atomic relaxations are given in Table I. In the zinc-blende structure there are two interstitial sites with tetrahedral symmetry; one in which the interstitial atom is surrounded by four cations, and the other where it is surrounded by four anions. In Fig. 1 we only show the interstitial anion (cation) with cation (anion) nearest neighbors, since we found that these are the lower-energy defects. In the next subsections the associated defect electronic structure, formation energies, and atomic structure are discussed.

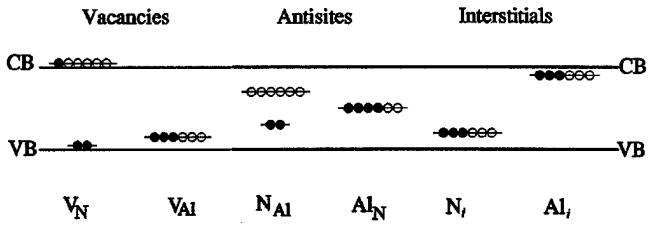


FIG. 2. Schematic representation of the defect-induced electronic states in zinc-blende AlN for the neutral charge state. Filled and open circles denote electrons and holes, respectively.

A. Electronic structure

In Fig. 2 the defect-induced Kohn-Sham levels are shown schematically for the neutral charge state in *zinc-blende* AlN. According to the present calculations, the nitrogen vacancy (V_N) introduces two states, an a_1 (s-like) state occupied by two electrons appearing above the valence-band maximum, and a triplet state t_2 (p-like) containing one electron, which is a resonance in the conduction band; the nitrogen vacancy thus acts as a single donor. (We keep in mind here that in the zinc-blende (ZB) structure, AlN has an indirect band gap, as noted earlier.) In the calculations, we find that the one electron associated with the t_2 state is located at the conduction-band edge and *not* in the defect level. Our calculated formation energy for the *neutral* charge state would therefore be unreliable (i.e., lower than it should be). Furthermore, within the DFT-LDA the band gap is notably smaller than experiment; corrections to the DFT-LDA would give rise to a larger band gap and also raise the formation energy of V_N^0. We can roughly estimate the formation energy of the nitrogen vacancy in the neutral charge state by approximating it to be the formation energy of the *positive* charge state when the Fermi level is at the bottom of the conduction band.[39] This

TABLE I. Defect geometries as a function of charge state for point defects and impurities in zinc-blende AlN. Displacements of the nearest-neighbor atoms are expressed as a percentage of the ideal bulk bond length. Positive values represent displacement away from the defect center, and negative values that toward it. For the N_{Al}^{EL2} defect, the values correspond to the percentage displacement of the N atom along the [111] direction with respect to the ideal N-Al distance, and for the N_i^{sp} defect, the values correspond to the percentage deviation in the distance between the two N atoms that share the substitutional site, referenced to the bond length of the N_2 molecule.

Charge state	V_N	V_{Al}	N_{Al}	N_{Al}^{EL2}	Al_N	N_i	N_i^{sp}	Al_i	Mg_{Al}	O_N	Si_{Al}
−3		9.2									
−2		9.7		38.8							
−1		9.7	7.5	45.2	17.3	−2.1	35.3		9.4		
0		10.5	0.9	48.9	19.0	0.5	24.8	−3.6	10.1		
1	2.3		8.2		21.0	3.1	15.4	−3.6		3.8	−5.0
2	9.1		15.4		23.4		9.3	−4.0			
3	17.3				25.9		4.8	−4.1			
4					28.5						

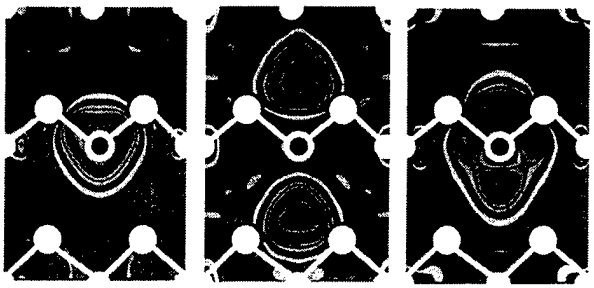


FIG. 3. Electron density distributions in the (110) plane of zinc-blende AlN for (left) the lower a_1-like nitrogen vacancy state, (center) the state at the conduction band minimum, and (right) the higher-lying t_2-like nitrogen vacancy state. Large and small white circles represent Al and N atoms, respectively, and the black dots indicate the nitrogen vacancy.

gives a value of about 7.8 eV for N-rich conditions, and one of about 4.3 eV for Al-rich conditions using the *GW* band gap[38] of 5.0 eV. We point out that attempts to occupy this upper defect level, i.e., considering negatively charged states (e.g., V_N^{1-}, V_N^{2-}, V_N^{3-}, etc.) will be unsuccessful, since, rather than the defect state, the conduction band will be occupied, leading to erroneous results. We believe that this was the case in Ref. 10.

The behavior of the nitrogen vacancy in *wurtzite* AlN is quite different. The band gap of wurtzite AlN is significantly larger than that of ZB AlN, and the upper defect state now lies *below* the conduction-band minimum. In order to confirm these assignments, we have performed calculations using self-interaction and relaxation-corrected (SIRC) pseudopotentials which afford an accurate description of the band gap:[40] for zinc-blende AlN we obtain a band gap of 5.05 eV (in very close agreement with the *GW* result) and for the wurtzite structure one of 6.50 eV. We have found that the t_2 nitrogen vacancy state is indeed *above* the conduction-band minimum in the zinc-blende structure and *below* it in the wurtzite structure. The calculations were carried out analogously to those reported in Ref. 33; further details will be given in a forthcoming publication.[41] In Ref. 33 we compared the position and nature of native defects in InN as obtained using the LDA and SIRC potentials, and found that the SIRC calculations affect the positions of some of the defect states in the band gap, but the general conclusions obtained from the standard DFT calculations remained valid. In Fig. 3, we show the spatial distribution (as obtained from our DFT-LDA calculations) of the (lower) a_1-like nitrogen defect state; the next highest state, which is an extended state at the conduction-band minimum (CBM); and the t_2-like nitrogen defect state, which is at an energy higher than the CBM state. The former and latter are clearly localized at the vacancy, while that at the CBM is a bulk state of Al character.

The aluminum vacancy (V_{Al}) creates a triplet state in the lower part of the band gap (see Fig. 2) which is occupied by three electrons and can be filled with three more electrons, i.e., V_{Al} is a triple acceptor. For the substitutional nitrogen antisite (N_{Al}) a doubly occupied singlet state in the band gap is introduced, as well as an empty triplet state higher in the

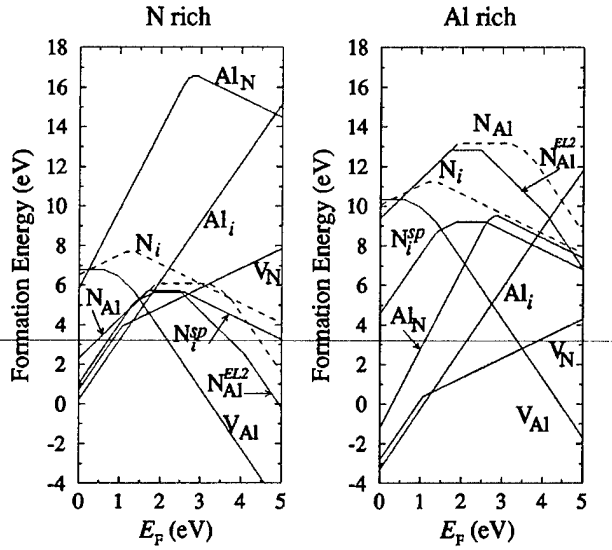


FIG. 4. Defect formation energies (LDA) for native point defects in zinc-blende AlN as a function of the Fermi level under nitrogen-rich (left panel) and aluminum-rich (right panel) conditions. $E_F=0$ corresponds to the valence-band maximum.

band gap which can trap six electrons. Depending on the position of the Fermi level, N_{Al} may act as an acceptor or a donor (i.e., this defect, and the following two discussed, exhibit an amphoteric behavior). The aluminum antisite (Al_N) introduces a deep triplet level deep in the band gap. It is occupied by four electrons, and can be filled with two more electrons, or emptied of all four, thus acting either as an acceptor or a donor. For the tetrahedrally coordinated nitrogen interstitial (N_i), a triplet state in the band gap results, which is occupied by three electrons. This level can be filled by three more electrons or conversely emptied of three electrons, again acting respectively as an acceptor or donor depending on the position of the Fermi level. The aluminum interstitial (Al_i), finally, introduces a triplet state in the upper part of the band gap, occupied by three electrons (see Fig. 2), and is thus a triple donor. For the Al interstitial at the *O* site (see Fig. 5) in the wurtzite structure, which is lower in energy than the *T* site (as discussed below), there is a doubly occupied single deep donor level and a higher lying singlet level occupied by one electron.

B. Formation energy

In the left and right panels of Fig. 4 we present the calculated defect formation energies for the various charge states as a function of the Fermi energy E_F in zinc-blende AlN, for N- and Al-rich conditions, respectively. For clarity, for each charge state of each defect we have not plotted the full straight line, but show only the section where that charge state is lower in energy than all other charge states. Thus the change in slope of the lines represents a change in the charge state of the defect [cf. Eq. (1)]. We note that the difference in formation energies of the vacancies between N- and Al-rich conditions is exactly the heat of formation of AlN: $\Delta H = E_{AlN}^{tot} - E_{Al,bulk}^{tot} - 1/2E_{N_2}^{tot}$, where E_{AlN}^{tot}, $E_{Al,bulk}^{tot}$, and $E_{N_2}^{tot}$ are

the total energies of bulk AlN, bulk Al, and the N_2 dimer, respectively. This is also the case for the interstitial defects, but here the difference is $-\Delta H$. The difference in formation energy between N- and Al-rich conditions for the antisites corresponds to twice the heat of formation. We calculate the heat of formation of AlN to be -3.40 and -3.05 eV for the LDA and GGA, respectively. The experimental value of ΔH is -3.28 eV.[25] Note that the negative formation energies occurring in Fig. 4 are not a problem; indeed, charge neutrality will constrain the Fermi level to a region where the formation energies are positive.

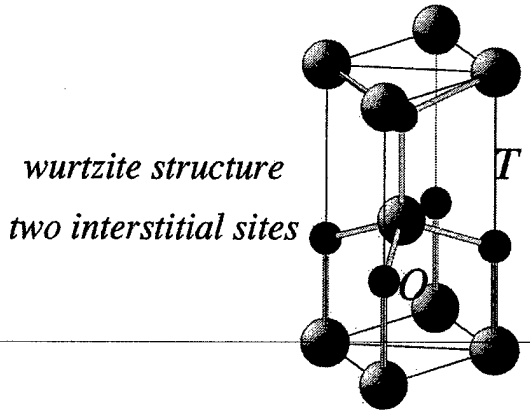


FIG. 5. Schematic representation of the wurtzite structure indicating the two high-symmetry interstitial positions T and O. Large gray and small dark spheres represent cations and anions, respectively.

1. Vacancies

It can be seen that under p-type conditions (low E_F values) the nitrogen vacancy in the triply positively charged state, V_N^{3+}, has a very low formation energy. We find that the doubly positively charged nitrogen vacancy V_N^{2+} is unstable, thus presenting a negative-U effect between triply and singly positive charge states. This has also been found to occur in GaN.[16,42] The transition level $E^{3+/+}$ of V_N in AlN is about 1.0 eV, to be compared with ≈ 0.2 eV in GaN.[42] Our results for V_N^{+} and V_N^{2+} are in good agreement with the calculations of Mattila and Nieminen;[10] those authors, however, did not investigate V_N^{3+}. In positive charge states (cf. our discussion in Sec. III A), we expect the calculated formation energies to be accurate, since the one electron that was a resonance in the conduction band is removed.[43] Under n-type conditions we see that the formation energy of the nitrogen vacancy is high, indicating that this defect will not occur in high concentrations. Nitrogen vacancies were previously shown not to be a cause of n-type conductivity in GaN (Ref. 44); combined with the present results for AlN, we can conclude that nitrogen vacancies will also not be a cause of n-type conductivity in $Al_xGa_{1-x}N$ alloys.

Under n-type conditions (large E_F values), the triply negatively charged aluminum vacancy V_{Al}^{3-} has the lowest formation energy. The behavior of the Al vacancy is similar to the case of the gallium vacancy in GaN. However, because of the larger band gap of AlN, the formation energy of V_{Al}^{3-} is significantly lower than V_{Ga}^{3-}. The defect level of V_{Al} lies close to the valence-band maximum, and is made up of states that reflect nitrogen-p (i.e., valence-band) character; we therefore expect that the corrections to the DFT-LDA band gap will not significantly affect the energetic position of this level, and consequently that the calculated formation energy of V_{Al}^{3-} will also be valid for all E_F values shown in Fig. 4.

Regarding the comparison between wurtzite and zinc-blende structures for V_N^{3+} and V_{Al}^{3-}, we found that the formation energies differ by only small amounts (by 0.01 and 0.13 eV, respectively). The defect geometries are also very similar, as discussed below. The main and important difference, as noted above, is that in a wurtzite structure the nitrogen vacancy induces an upper defect level below the CBM, which in the neutral charge state is occupied by a single electron, while in a zinc-blende structure this state is a resonance in the conduction band.

2. Interstitials

The defect level induced by the aluminum interstitial, Al_i, lies high in the band gap, close to (or at) the conduction-band minimum. On emptying this state the formation energy is rapidly lowered as the Fermi level moves down. It can be seen that, for p-type conditions, the formation energy is very low, comparable to V_N^{3+}. In the 3+ charge state the Al_i level has been emptied of all electrons; hence we anticipate that the calculated formation energy is accurate, using the same argument made above for the positive charge state of the nitrogen vacancy.

The equilibrium geometry for the Al-interstitial defect in zinc-blende AlN discussed above is that of the symmetric tetrahedral interstitial site, i.e., there is no tendency for the interstitial atom to assume a lower symmetry position. This is in contrast to the Al interstitial site in the wurtzite structure, which we also studied; in a wurtzite structure there are also two types of high-symmetry interstitial positions: (tetrahedral) T and (octahedral) O sites. The T site corresponds to the center of the line joining the closest nonbonded Al and N atoms along the c axis (see Fig. 5). It has two nearest neighbors (one cation and one anion) and six next-nearest neighbors (three cations and three anions). The O site is located at the center of the hexagonal channel, and has six nearest neighbors (three cations and three anions). For our wurtzite calculations, when using a 32-atom supercell, with six $\mathbf{k}$ points in the Brillouin zone and imposing no symmetry restrictions, we found that the O site is lower in energy than the T site by 1.53 eV, the formation energies for the O site being 14.30 and 10.76 eV for the neutral charge state under N- and Al-rich conditions, respectively. These values are more than 2.0 eV *higher* than for Al_i in the zinc-blende structure. The Al interstitial is thus not expected to be an important defect in wurtzite material. This difference is due to the different local atomic geometry of the interstitial sites in the two structures: in the zinc-blende structure the interstitial Al atom can be tetrahedrally bonded to *four* N atoms, but in wurtzite material, only three Al_i-N bonds are formed. A similar difference between wurtzite and zinc-blende struc-

tures was recently reported for interstitial beryllium impurities in GaN, and also explained in terms of a difference in coordination.[45]

In GaN it has been shown that a "(100) split-interstitial" configuration has the lowest formation energy[46] for an interstitial N atom. In this geometry two N atoms share the same substitutional site with a bond length similar to the N_2 dimer, and each N atom in the defect has two bonds to the surrounding cation atom instead of four. In AlN, we find that this N split-interstitial geometry, N_i^{sp}, is also energetically more favorable than the tetrahedrally coordinated interstitial site (see Fig. 4): in the neutral charge state its energy is lower by $\approx$2.0 eV. The (100) N split-interstitial defect introduces a pair of deep levels very close in energy which are occupied by three electrons. In Table I the percentage deviation in the bond length of the two N atoms sharing the same substitutional site is given with respect to that of the free N_2 molecule.

3. Antisites

Mattila *et al.*[11] identified the stability of an $EL2$ geometry for the N antisite in the negative charge states in both zinc-blende (ZB) and wurtzite (WZ) GaN and AlN. In this atomic configuration, there is a displacement of the nitrogen atom off the substitutional position along the [111](ZB) and [0001](WZ) directions. We have performed calculations for the N antisite in AlN investigating the stability of the $EL2$ geometry; we find that even in the neutral charge state the formation energy is lower than the substitutional antisite by 0.35 eV (see Fig. 4). In the neutral $EL2$ configuration there is a doubly occupied singlet state and above this level there are two deep unoccupied singlet states. We find that the singly negative charge state is not stable at any position of the Fermi level, therefore exhibiting a negative-U behavior. This was also found to be the case for GaN.[11] The displacement of the N atom along the [111] direction for the neutral charge state is 49% of the ideal Al-N nearest-neighbor distance, in agreement with Ref. 11. The displacements for the singly and doubly negative charge states are given in Table I. While the behavior of N_{Al}^{EL2} is very interesting, we doubt that this defect will play an important role in the electronic properties of AlN; indeed, Fig. 4 shows that aluminum vacancies are always lower in energy than nitrogen antisites in n-type AlN, and will therefore be the preferred compensating center.

Under N-rich conditions the Al antisite is a high energy defect. Under Al-rich conditions it has a low formation energy in the 4^+ charge state for strongly p-type material; it is however, higher than the V_N^{3+} and Al_i^{3+} defects and thus not expected to play an important role.

C. Comparison of GGA and LDA results

Figure 6 shows the comparison between formation energies obtained using the GGA vs the LDA. Only the lowest-energy defects and charge states (V_N^{3+}, Al_i^{3-}, and V_{Al}^{3-}) were included in this study. It can be seen that there is good qualitative agreement but that some quantitative deviations

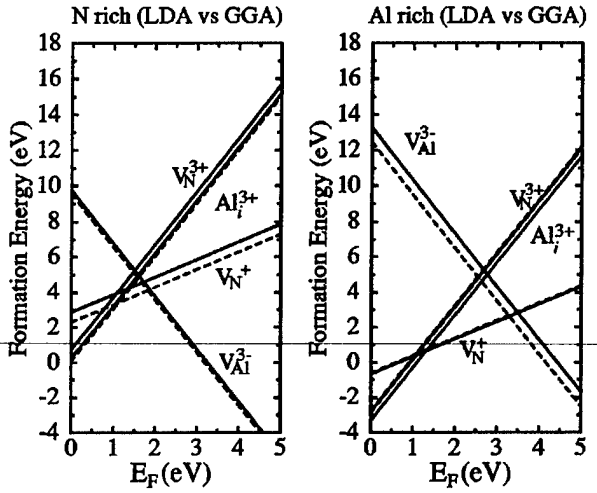


FIG. 6. Comparison of GGA (dashed lines) and LDA (solid lines) results for defect formation energies of various low-energy point defects in zinc-blende AlN. Formation energies are shown as a function of the Fermi level under nitrogen-rich (left panel) and aluminum-rich (right panel) conditions. E_F=0 corresponds to the valence-band maximum.

occur. We were able to attribute the differences to two main effects: (i) Differences in the energies of the reservoirs [chemical potentials; see Eq. (1)] due to the LDA/GGA differences in cohesive energies, binding energies, and heats of formation. In particular, the cohesive energy of bulk Al is 0.65 eV smaller in the GGA than in the LDA; the binding energy of the N_2 molecule is 0.86 eV smaller (per N atom) in the GGA; and the magnitude of the heat of formation of bulk AlN is 0.35 eV smaller in the GGA.[31] We note that the GGA results for each of these systems are in better agreement with experiment (see Ref. 31). The LDA overbinding in the defect calculations is to some degree compensated due to the fact that the reservoir or reference systems also exhibit an overbinding; however, a complete cancellation does not occur. (ii) Differences in the band gap. The LDA and GGA band structures are extremely similar when calculated at the same lattice constant. For instance, at the *experimental* lattice constant the LDA and GGA band gaps differ by less than 0.02 eV. However, the point-defect calculations are carried out at the respective *theoretical lattice constants*, at which we find that the GGA band gap is 0.62 eV smaller than the LDA band gap.[31] The reduction in the band gap causes some of the defect-induced levels to lie at lower energies in the band gap, which could affect the calculated formation energy of defects for which these levels are occupied.

D. Atomic structure

In Table I we report the change in defect geometry with charge state for point defects in zinc-blende AlN. The displacement of the nearest-neighbor atoms of the defect with respect to the unrelaxed positions is given as a percentage of the ideal Al-N bond length. The nitrogen vacancy exhibits a strong dependence on the charge state: In the neutral charge state, the four surrounding Al atoms are almost unperturbed,

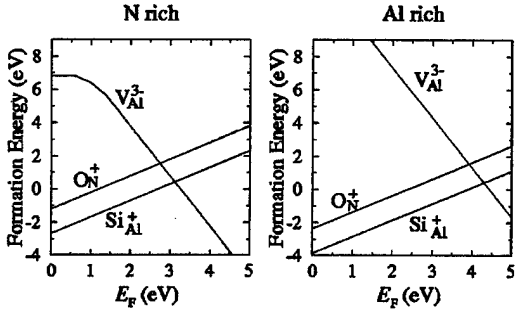


FIG. 7. Formation energies (LDA) of impurities and defects relevant for n-type AlN, as a function of the Fermi level under nitrogen-rich (left panel) and aluminum-rich (right panel) conditions. Oxygen and silicon are donor impurities, and the Al vacancy is the dominant compensating defect. $E_F=0$ corresponds to the valence-band maximum.

while, for an increasing positive charge state, the Al atoms move increasingly outwards. For the 3+ charge state, the displacement is 17%. (The corresponding GGA value is about 14%.) This can be compared to that of 19% for the same defect in GaN.[42] In a wurtzite structure, we find that the three equivalent nearest-neighbor Al atoms move outwards from the vacancy by 18% and the nonequivalent neighboring Al atom moves outwards by 27%.

In contrast, relaxations around the aluminum vacancy show very little dependence on the charge state, with roughly a 9–10 % outward breathing relaxation of the four neighboring N atoms. (The corresponding GGA value for V_{Al}^{3-} is 9%.) For V_{Al}^{3-} in wurtzite AlN the nearest-neighbor N atom along the c axis moves outward by $\approx 15\%$, and the three equivalent neighboring N atoms move outwards by 9%, similar to the relaxations in the zinc-blende structure.

The geometry of the aluminum interstitial, Al_i, also does not depend very strongly on the charge state, exhibiting a small *inward* movement of the neighboring N atoms of about 4%. The nitrogen split interstitial is the energetically favorable configuration for the N interstitial. It can be seen that for the 3+ charge states the distance between the two N atoms is only about 5% larger than in the N_2 molecule. When electrons are added to the defect level, the N-N distance is increased. Large outward relaxations are found for the aluminum antisite; this is due to the larger size of the Al atom as compared to the N atoms, and induces large internal strains, leading to rather high formation energies. The relaxations of Al_N become larger for increasing positive charge states, from 17% for Al_N^- to 28% for Al_N^{4+}. The substitutional nitrogen antisite N_{Al} induces only very small displacements for the neutral charge state (1%), and increasingly larger outward displacements for *both* negative and positive charge states. We saw above that the energetically favorable configuration for N_{Al} is the so-called $EL2$ geometry. In this configuration the N atom has moved between 49% (neutral) and 39% (doubly negative) of the ideal Al-N distance along the [111] direction, thus forming shorter and stronger bonds with its three neighboring N atoms (as opposed to a tetrahedral bonding arrangement).

IV. IMPURITIES

A. Mg, O, and Si impurities

Figure 7 shows the calculated formation energy in zinc-blende AlN of the O and Si impurities, as well as V_{Al}^{3-}, as a function of E_F for both N- and Al-rich conditions. As found in earlier studies of O and Si in GaN (Ref. 44) and AlN,[11] oxygen substitutes on the N site and silicon substitutes on the cation site. The O and Si impurities introduce a singly occupied level which is a resonance in the conduction band. From Table I it can be seen that O_N^+ induces an outward expansion of the nearest four Al atoms of 3.8%, while for Si_{Al}^+ an inward relaxation of 5.0% of the neighboring N atoms occurs. The formation energies of O and Si in the positive charge state are very low in p-type material.

Magnesium in GaN has been shown to have the lowest formation energy when occupying the substitutional cation site.[47] We therefore assumed that for AlN it would also "prefer" the cation site and act as an acceptor. The Mg atom induces a large outward movement of the surrounding N atoms of 10.1% for Mg^0 and 9.4% for Mg^-. Figure 8 displays the calculated formation energy of the Mg impurity as well as of the dominant compensating defects in p-type AlN, namely, $V_N^{3+/+}$ and Al_i^{3-}. The lowest-energy charge state of Mg changes from neutral to -1 at $E_F=0.4$ eV; this transition level corresponds to the acceptor ionization energy. This value can be compared to that calculated for GaN of ≈ 0.2 eV.[47] Experimentally, based on an analysis of temperature-dependent Hall data, Tanaka *et al.*[48] found the Mg activation energy to be 35 meV deeper in $Al_{0.08}Ga_{0.92}N$ than in GaN. Assuming linearity, this would make the Mg acceptor in AlN 0.44 eV deeper than in GaN, in reasonable qualitative agreement with our computed result.

The low formation energies found in the present work for the triple donors, V_N^{3+} and Al_i^{3+}, under p-type conditions, indicate that these defects will compensate shallow acceptors (e.g., Mg), thus potentially interfering with attempts at p-type doping. In particular, the nitrogen vacancy has a lower formation energy in AlN than in GaN, indicating that compensation by nitrogen vacancies is therefore the likely cause of the decreased doping efficiency of Mg when the Al content is raised in $Al_xGa_{1-x}N$ alloys. The Al interstitial should not be a concern in wurtzite material, but is low in energy in zinc-blende AlN. Problems in achieving p-type $Al_xGa_{1-x}N$ for $x \geq 0.13$ have indeed been observed.[49] The low formation energy of the O and Si impurities in the positive charge state imply that either of these impurities, if present during p-type growth, will be readily incorporated and act as a compensating center.

We find that V_{Al}^{3-} has a sufficiently low formation energy to act as an effective compensating center for shallow donors (e.g., O and Si), inhibiting successful n-type doping. Severe problems in achieving n-type conductivity were reported in $Al_xGa_{1-x}N$ for $x \geq 0.4$; see, for example, Refs. 49–51. In Ref. 12 we proposed that the drop in n-type conductivity with increasing x can be attributed to two effects: (i) In the case of doping with oxygen (the most common unintentional donor), a DX transition occurs (in wurtzite material) which

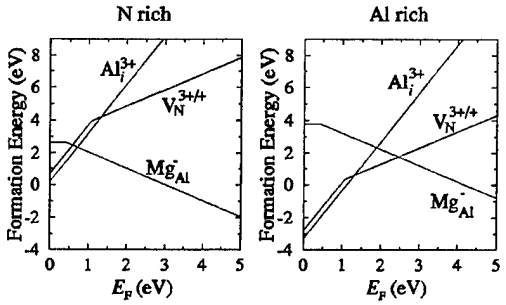


FIG. 8. Formation energies (LDA) of impurities and defects relevant for p-type AlN, as a function of the Fermi level under nitrogen-rich (left panel) and aluminum-rich (right panel) conditions. Magnesium is the acceptor impurity, and Al_i and V_N are the dominant compensating defects. $E_F=0$ corresponds to the valence-band maximum.

converts the shallow donor into a deep level;[43] the DX behavior of oxygen is not included in Fig. 7. (ii) Compensation by the cation vacancies (V_{Ga} or V_{Al}, triple acceptors) increases with alloy composition x.

B. Mg_nO_n complexes

We also performed calculations for complex formation between Mg and O in zinc-blende AlN, in particular MgO and Mg_2O_2. We consider Mg substituting on Al sites and O atoms on N sites. We calculate (i) the formation energy as given in Eq. (1) and (ii) the average binding energy, defined with respect to the separated impurities Mg_{Al}^- and O_N^+:

$$E^b(nMg_{Al};nO_N)=[-E^f(nMg_{Al};nO_N)+nE^f(Mg_{Al}^-) +nE^f(O_N^+)]/n, \quad (2)$$

where E^f are the formation energies of the defects indicated in the associated parentheses, and $n=1$ or 2.

The atomic structure of the MgO and Mg_2O_2 defects (at the ideal, unrelaxed positions) is shown in Fig. 9, where the binding energies are also given. The relaxed Mg-O bond lengths remain very similar in each Mg_nO_n defect, namely 2.09 Å for MgO and 2.08 Å for Mg_2O_2. These values are very close to those of crystalline MgO, where the bond length is 2.10 Å.[36] Compared to the N-Al nearest-neighbor distance (1.87 Å), the Mg-O bond lengths are about 11.5% longer. As discussed above, the Mg_{Al} defect is a shallow acceptor, and the O_N defect is a shallow donor. On forming the $Mg_{Al}O_N$ defect, we find that there are no longer any states in the band gap; there is an occupied defect state below the VBM and an unoccupied state above the CBM. The neutral MgO complex has a binding energy of 0.6 eV, as reported previously.[20] For the Mg_2O_2 defect, which forms an O-Mg-O-Mg zigzag structure, the binding energy per MgO unit is somewhat larger, namely, 0.8 eV. Again, in this case there are no states in the band gap. The stability of the com-

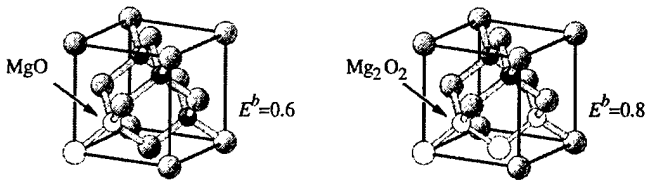


FIG. 9. Atomic geometry of MgO and Mg_2O_2 complexes in AlN. Large and small open circles represent Mg and O atoms, respectively, and large pale gray and small dark gray circles represent Al and N atoms, respectively. E^b is the binding energy (in eV) per MgO pair.

plexes can be understood as reflecting the larger bond strength of Mg-O as compared to Mg-N: the heat of formation of crystalline MgO is -6.16 eV, and for MgO_2 the value is -6.46 eV, as compared to -4.81 eV for Mg_3N_2.[36] The formation energies of the MgO and Mg_2O_2 defects are 0.92 and 1.67 eV, respectively. These quantities are independent of E_F, since the complexes are neutral, and are the same for N- and Al-rich conditions due to the assumed chemical potentials described above. Thus, on thermodynamic grounds, due to its relatively high formation energy, we conclude that the Mg_2O_2 complex is unlikely to be present in any significant quantities. MgO pairs, however, may be more abundant and could play a role in the compensation of acceptor doping with Mg if oxygen is present as a contaminant.

V. CONCLUSION

We have performed first-principles calculations for native point defects and selected impurities in AlN. The triply positively charged nitrogen vacancy (and the triply positively charged aluminum interstitial, in zinc-blende AlN) have the lowest formation energies in p-type material, and in n-type material the triply negatively charged aluminum vacancy has the lowest formation energy. Due to the low formation energies, the triple donors V_N^{3+} and Al_i^{3+} may act as compensating centers when trying to achieve p-type doping (e.g., with Mg). If oxygen contaminants are present in the material, then substitutional $Mg_{Al}O_N$ pairs may also contribute to the compensation. In n-type material, V_{Al}^{3-} may act as a compensating center, thus limiting the electron concentration provided by Si and O impurities which we find are shallow donors. The formation energy of the nitrogen vacancy is high in n-type AlN, indicating that this defect is not expected to occur in high concentrations. V_N^{2+} is found to be unstable; thus there is a negative-U effect between V_N^+ and V_N^{3+}. For the nitrogen antisite we find that the $EL2$ configuration is the stable one for the neutral and negative charge states, where the singly negative charge state is found to be unstable, so that there is also a negative-U effect between $N_{Al}^{EL,0}$ and $N_{Al}^{EL,2-}$. This defect, however, has a high formation energy, indicating that it will not be present in high concentrations. For the N interstitial, we identified the stability and energetic

preference for a (100) split interstitial configuration. Again it is a high-energy defect. We find that the nature of the nitrogen vacancy is different in the zinc-blende and wurtzite structures in that the higher lying defect-induced level is a resonance in the conduction band in a zinc-blende structure but lies well below the conduction-band edge in a wurtzite structure. The nitrogen vacancy is thus a shallow donor in a zinc-blende structure, but a deep donor in wurtzite AlN. From comparison of the GGA and LDA results, for the systems investigated, we find some quantitative differences but *qualitative* agreement.

ACKNOWLEDGMENTS

This work was supported in part by the Air Force Office of Scientific Research, Contract No. F4920-00-C-0019. C.S. gratefully acknowledges support from the DFG (Deutsche Forschungsgemeinschaft).

[1] H. Morkoç, S. Strite, G. B. Gao, M. E. Lin, B. Sverdlov, and M. Burns, J. Appl. Phys. **76**, 1363 (1994), and references therein.

[2] F. A. Ponce and D. P. Bour, Nature (London) **386**, 351 (1997), and references therein.

[3] S. Nakamura, Solid State Commun. **102**, 237 (1997), and references therein.

[4] S. Nakamura, T. Mukai, and M. Senoh, Appl. Phys. Lett. **64**, 1687 (1994); S. Nakamura, M. Senoh, N. Iwasa, S. Nagahama, T. Yamada, and T. Makai, Jpn. J. Appl. Phys. **34**, L1332 (1995).

[5] S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, H. Kiyaku, and Y. Sugimoto, Jpn. J. Appl. Phys. **35**, L74 (1996); S. Nakamura, in *III-V Nitrides*, edited by F. A. Ponce, T. V. Moustakas, I. Akasaui, and B. A. Monemar, Mater. Res. Soc. Symp. Proc. No. 449 (Materials Research Society, Pittsburgh, 1996), p. L74.

[6] K. Itaya, M. Onomura, J. Nishio, L. Sugiura, S. Saito, M. Suzuki, J. Rennie, S. Nunoue, M. Yamamoto, H. Fujimoto, Y. Kokubun, Y. Ohba, G. Hatakoshi, and M. Ishikawa, Jpn. J. Appl. Phys. **35**, L1317 (1996).

[7] U. Kaufmann, M. Kunzer, M. Maier, H. Obloh, A. Ramakrishnan, B. Santic, and P. Schlotter, Appl. Phys. Lett. **72**, 1326 (1998).

[8] J. I. Pankove, J. T. Torvik, C.-H. Qiu, I. Grzegory, S. Porowski, P. Quigley, and B. Martin, Appl. Phys. Lett. **74**, 416 (1999).

[9] D. W. Jenkins and J. D. Dow, Phys. Rev. B **39**, 3317 (1989).

[10] T. Mattila and R. M. Nieminen, Phys. Rev. B **54**, 16 676 (1996); **55**, 9571 (1997).

[11] T. Mattila, A. P. Seitsonen, and R. M. Nieminen, Phys. Rev. B **54**, 1474 (1996).

[12] C. Stampfl and C. G. Van de Walle, Appl. Phys. Lett. **72**, 459 (1998).

[13] C. Stampfl and C. G. Van de Walle, in *Nitride Semiconductors*, edited by F. A. Ponce, S. P. Den Baars, B. K. Meyer, S. Nakamura, and S. Strite, MRS Symposia Proceedings No. 482 (Materials Research Society, Pittsburgh, 1998), p. 905.

[14] A. Fara, F. Bernardini, and V. Fiorentini, J. Appl. Phys. **85**, 2001 (1999).

[15] P. Boguslawski and J. Bernholc, Phys. Rev. B **56**, 9496 (1997).

[16] C. H. Park and D. J. Chadi, Phys. Rev. B **55**, 12 995 (1997).

[17] I. Gorczyca, A. Svane, and N. E. Christensen, Solid State Commun. **101**, 747 (1997).

[18] I. Gorczyca, A. Svane, and N. E. Christensen, Phys. Rev. B **60**, 8147 (1999); Acta Phys. Pol. A **92**, 785 (1997).

[19] I. Gorczyca, A. Svane, and N. E. Christensen, Phys. Rev. B **61**, 7494 (2000).

[20] C. G. Van de Walle, C. Stampfl, and J. Neugebauer, J. Cryst. Growth **189/190**, 505 (1998).

[21] M. Bockstedte, A. Kley, J. Neugebauer, and M. Scheffler, Comput. Phys. Commun. **107**, 187 (1997).

[22] D. M. Ceperley and B. I. Alder, Phys. Rev. Lett. **45**, 566 (1980); P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[23] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).

[24] The GGA has been found to improve the description of total energies, ionization energies, and electron affinities of atoms, atomization energies of molecules, and some solid state properties (Ref. 23), in addition to the adsorption energies [see, e.g., P. H. T. Philipsen, G. te Velde, and E. J. Baerands, Chem. Phys. Lett. **226**, 583 (1994); P. Hu, D. A. King, S. Crampin, M.-H. Lee, and M. C. Payne, *ibid.* **230**, 501 (1994)]; and reaction energies [see, e.g., D. Porezag and M. R. Pederson, J. Chem. Phys. **102**, 9345 (1995); J. Baker, M. Muir, and J. Andzelm, *ibid.* **102**, 2063 (1995)]; Furthermore, the GGA has been shown to be crucial in obtaining activation energy barriers consistent with experiment for H_2 dissociation [see, e.g., B. Hammer, K. W. Jacobsen, and J. K. Norskov, Phys. Rev. Lett. **70**, 3971 (1993); B. Hammer, M. Scheffler, K. W. Jarcobsen, and J. K. Norskov, *ibid.* **73**, 1400 (1994)]; Also the relative stability of structural phases appear to be better described for magnetic [see, e.g., T. C. Leung, C. T. Chan, and B. N. Harmon, Phys. Rev. B **44**, 2923 (1991)] and nonmagnetic systems [see, e.g., N. Moll, M. Bockstedte, M. Fuchs, E. Pehlke, and M. Scheffler, *ibid.* **52**, 2550 (1995); D. R. Hamann, Phys. Rev. Lett. **76**, 660 (1996); We refer to Ref. 30 for a detailed discussion].

[25] *Properties of Group III-Nitrides*, edited by J. H. Edgar, EMIS Datareviews Series (INSPEC, London, 1994).

[26] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[27] S. L. Cunningham, Phys. Rev. B **10**, 4988 (1974).

[28] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).

[29] M. Fuchs and M. Scheffler, Comput. Phys. Commun. **119**, 67 (1999).

[30] M. Fuchs and M. Scheffler, Phys. Rev. B **57**, 2134 (1998).

[31] C. Stampfl and C. G. Van de Walle, Phys. Rev. B **59**, 5521 (1999).

[32] C. Stampfl and C. G. Van de Walle, Phys. Rev. B **57**, R15 052 (1998).

[33] C. Stampfl, C. G. Van de Walle, D. Vogel, P. Krüger, and J. Pollmann, Phys. Rev. B **61**, R7846 (2000).

[34] D. B. Laks, C. G. Van de Walle, P. E. Blöchl, and S. T. Pantelides, Phys. Rev. B **45**, 10 965 (1992).

[35] A. Garćia and J. E. Northrup, Phys. Rev. Lett. **74**, 1131 (1995).

[36] *Handbook of Chemistry and Physics*, 76th ed., edited by D. R. Lide (CRC Press, Boca Raton, FL, 1995).

[37] J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983); L. J. Sham and M. Schlüter, *ibid.* **51**, 1888 (1983).
[38] A. Rubio, J. L. Corkhill, M. L. Cohen, E. Shirley, and S. G. Louie, Phys. Rev. B **48**, 11 810 (1993).
[39] S. B. Zhang and D. J. Chadi, Phys. Rev. B **42**, 7174 (1990).
[40] D. Vogel, P. Krüger, and J. Pollmann, Phys. Rev. B **54**, 5495 (1996).
[41] C. Stampfl, C. G. Van de Walle, M. Fuchs, D. Vogel, P. Krüger, and J. Pollmann (unpublished).
[42] J. Neugebauer and C. G. Van de Walle, Festköerperprobleme **35**, 25 (1996).
[43] C. G. Van de Walle, Phys. Rev. B **57**, R2033 (1998).
[44] J. Neugebauer and C. G. Van de Walle, in *22nd International Conference on the Physics of Semiconductors, Vancouver, Canada 1994* (World Scientific, Singapore 1995), p. 2327.
[45] C. G. Van de Walle, S. Limpijumnong, and J. Neugebauer, Phys. Rev. B **63**, 245205 (2001).
[46] J. Neugebauer and C. G. Van de Walle, Phys. Rev. B **50**, 8067 (1994).
[47] J. Neugebauer and C. G. Van de Walle, Appl. Phys. Lett. **68**, 1829 (1996).
[48] T. Tanaka, A. Watanabe, H. Amano, Y. Kobayashi, I. Akasaki, S. Yamazaki, and M. Koike, Appl. Phys. Lett. **65**, 593 (1994).
[49] M. D. Bremser, W. G. Perry, T. Zheleva, N. V. Edwards, O. H. Nam, N. Parikh, D. E. Aspnes, and R. F. Davis, MRS Internet J. Nitride Semicond. Res. **1**, 8 (1998).
[50] H. G. Lee, M. Gershenzon, and B. L. Goldenberg, J. Electron. Mater. **20**, 621 (1991).
[51] X. Zhang, P. Kung, A. Saxler, D. Walker, T. C. Wang, and M. Razeghi, Appl. Phys. Lett. **67**, 1745 (1995).

Entropy-Driven Stabilization of a Novel Configuration for Acceptor-Hydrogen Complexes in GaN

Sukit Limpijumnong,* John E. Northrup, and Chris G. Van de Walle
Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, California 94304
(Received 13 July 2001; published 29 October 2001)

We present a model for the microscopic structure of Mg-H complexes in GaN, explaining the unusual bond angle observed in recent vibrational spectroscopy studies. The structure is not the lowest-energy configuration at $T = 0$, but it is stabilized at elevated temperatures due to the large entropy associated with a set of low-energy rotational excitations. The rotational excitation spectrum is calculated using a quantum-mechanical model in which the hydrogen atom moves in a weak corrugation potential. Consequences for experiment are discussed.

DOI: 10.1103/PhysRevLett.87.205505 PACS numbers: 61.72.Bb, 61.72.Vv, 63.20.Pw

It is well known that hydrogen can passivate dopants in semiconductors, resulting in the formation of electrically inactive dopant-hydrogen complexes [1]. This process plays a critical role during the growth of acceptor-doped GaN, where incorporation of H results in complete neutralization of the acceptors. A post-growth anneal is required to remove H from the vicinity of the acceptors, yielding the desired p-type conductivity. A thorough understanding of complex formation and dissociation is deemed essential for optimizing the activation process and enhancing the conductivity. Experimentally, vibrational spectroscopy has proven an excellent tool for monitoring the presence of acceptor-hydrogen complexes [2]. Identification of the complex responsible for a particular line, however, is possible only in conjunction with first-principles theory. For instance, the local vibrational mode (LVM) at 3125 cm^{-1} observed in Mg-doped GaN [3,4] occurs at a much higher frequency than expected for Mg-H bonds. Computations have shown, however, that the hydrogen atom in the Mg-H complex is strongly bonded to a nitrogen neighbor of the Mg atom [see Fig. 1(a)], explaining the high frequency of the observed vibrational mode [5].

Recently, infrared spectroscopy using polarized light produced additional information about the Mg-H complex: Clerjaud *et al.* [4] found that the electric dipole induced by the vibration (which should roughly correspond to the orientation of the N-H bond) formed an angle of 130° with the c axis of the GaN wurtzite crystal. This large deviation from the expected angle of 109° in the $AB_{N,\perp}$ (antibonding) configuration [see Fig. 1(a)] is very puzzling. Our calculations show that a distorted $AB_{N,\perp}$ configuration in which the N-H bond forms an angle of 130° with the c axis (as shown in Ref. [4]) is unstable and immediately relaxes back to the regular configuration with $\theta = 109°$. The first goal of this Letter is to propose a new model for the configuration of the Mg-H complex, illustrated in Fig. 1(b): in this structure (which we label $OA_{\parallel}$), the N-H bond is *not* aligned in the same direction as the Mg-N bond [as in Fig. 1(a)]; rather, the Mg-N bond is oriented (roughly) along [0001], and the interaction between H and the Mg atom changes the angle θ to approximately 130°. Both this angle and the calculated vibrational frequency for this structure (reported below) are in very good agreement with experiment [4]. However, one mystery remains: the calculated total energy of this configuration is *higher* (by 0.19 eV) than the energy of the $AB_{N,\perp}$ configuration. The magnitude of this energy difference would render experimental observation of $OA_{\parallel}$ very unlikely.

The second goal of this Letter is to provide an explanation for the stabilization of the 130° configuration. The relative occupation of the two configurations is determined at high temperatures, where the H atom is sufficiently mobile to explore various local minima. The stability of the candidate configurations is thus determined by their *free energy*. It turns out the 130° configuration gives rise to a set of low-energy excitations, which contribute significantly to the *entropy* of the complex. At sufficiently high temperatures the entropy term lowers the free energy of the

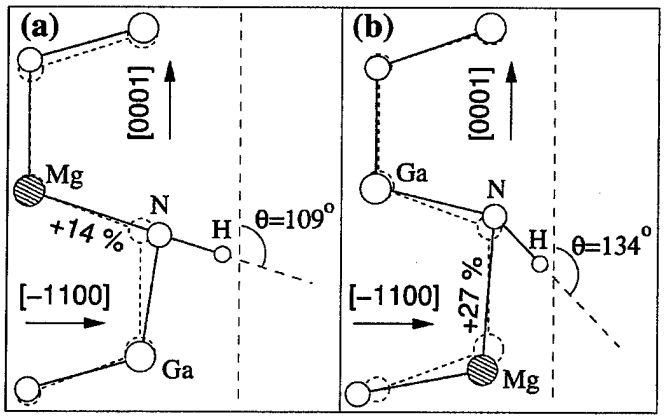


FIG. 1. Schematic representation of atomic positions in the (11-20) plane for a Mg-H complex in wurtzite GaN, with H at the (a) $AB_{N,\perp}$ site and (b) $OA_{\parallel}$ site. Large circles represent Ga atoms, medium circles N atoms, shaded circle the Mg atom, and the small circle the H atom. Dashed circles indicate ideal atomic positions, dashed lines bonds in the ideal lattice. Changes in Mg-N bond lengths are given as a percentage of change from the bulk Ga-N bond lengths.

 © 2001 The American Physical Society

$OA_{\parallel}$ configuration below that of $AB_{N,\perp}$. We will show quantitatively that this entropy-driven mechanism stabilizes the $OA_{\parallel}$ configuration at the temperatures relevant for the GaN:Mg-H system. Other consequences of the presence of low-energy excitations will be discussed.

Our results are based on first principles calculations using density functional theory in the local density approximation [6] and norm-conserving pseudopotentials, with a plane-wave basis set [7]. We have performed calculations both with explicit inclusion of the Ga $3d$ states (requiring a plane-wave cutoff of 60 Ry) and with the "nonlinear core correction" (nlcc) [8] (with a cutoff of 40 Ry); the results are not sensitive to this approximation, and the numbers reported below were obtained with the nlcc. A supercell geometry was employed, with supercells containing 32 or 64 atoms for zinc blende, and 32 or 96 atoms for wurtzite, with at least 24, 36, 24, and 46 atoms, respectively, allowed to relax. The larger supercells yield converged results for all the relevant energy differences, while the smaller cells are adequate for extracting the energy differences needed to calculate vibrational frequencies.

These frequencies are obtained by calculating energies for configurations in which the H is displaced from its equilibrium site, in accordance with a stretching- or bending-mode vibration. Because of the small mass of the H atom, displacements of the host atoms can be neglected; we have shown, by explicit comparison with evaluating the dynamical matrix [9], that this is a very good approximation, provided the reduced mass of the N-H system is used in evaluating the frequencies based on force constants. These constants are obtained from fitting a fourth-order polynomial to the calculated data points; the third- and fourth-order terms allow us to derive anharmonic corrections, which are sizable in the case of a light atom such as H [10]. The numerical error bars associated with this approach are fairly small, but a systematic error results from the fact that our pseudopotential-density-functional calculations overestimate the N-H bond length by 0.6%. We have checked the systematic nature of this deviation in a series of calculations on many different configurations of H in GaN as well as on NH_3 molecules [11]. The deviation results in an underestimation of the vibrational frequencies of stretch modes by 75 cm^{-1}.

The first step in our investigations is to establish all candidate configurations for the Mg-H complex. The AB_N and BC (bond center) configurations have previously been investigated for both the zinc-blende [5] and the wurtzite [12] structures, with the AB_N site showing lower energy. In the wurtzite structure, there are two types of each AB_N and BC: one type has threefold symmetry and is associated with bonds oriented parallel to the c axis; we label these sites $BC_{\parallel}$ and $AB_{N,\parallel}$. The other type is associated with the bonds that are *not* parallel to the c axis; although these are not exactly "perpendicular" to the c axis, we use the notation $BC_{\perp}$ and $AB_{N,\perp}$.

We find that the BC sites are not stable or even metastable sites for H. For instance, starting from $BC_{\parallel}$ and moving H slightly off the c axis causes this configuration to relax to the $OA_{\parallel}$ (off-axis) configuration illustrated in Fig. 1(b). This relaxation is caused by the large size of the Mg atom (covalent radius 1.36 Å) which squeezes the H out of the $BC_{\parallel}$ position to the $OA_{\parallel}$ position, where H is displaced by 0.79 Å from the ideal c-oriented Mg-N bond direction. The calculated angle between the N-H bond and the c axis is $\theta = 134°$. The size of the cation as the main driving force for the off-axis configuration is confirmed by comparison with other H-related centers in GaN: The small size of Be (covalent radius 0.90 Å) allows $BC_{\parallel}$ to remain most stable [9], while in pure GaN (Ga covalent radius 1.26 Å), only a very slight displacement (<0.15 Å) off the c axis occurs.

Results for energy differences and lattice relaxations for all relevant structures are listed in Table I. The $AB_{N,\perp}$ configuration [illustrated in Figs. 1(a) and 2(a)] has the lowest energy (which we set to zero in Table I). Next higher in energy is the $OA_{\parallel}$ configuration [illustrated in Figs. 1(b) and 2(b)]. The $OA_{\perp}$ configuration, which is the off-axis version of $BC_{\perp}$, has the highest energy. Within the accuracy of the calculations, $AB_{N,\perp}$ and $OA_{\parallel}$ have essentially the same N-H bond length and very similar vibrational frequencies for the stretch mode (3017 and 3034 cm^{-1}). As mentioned above, the calculated vibrational frequencies underestimate the real value by about 75 cm^{-1}; adding this correction brings the value for $OA_{\parallel}$ up to 3109 cm^{-1}, in very good agreement with the experimental value of 3125 cm^{-1}. While the vibrational frequency itself does not allow us to distinguish between $AB_{N,\perp}$ and $OA_{\parallel}$, the geometry of $OA_{\parallel}$, with the angle θ equal to 134°, is in excellent agreement with the polarized-light results of Clerjaud *et al.* [4], who derived $\theta = 130 \pm 5°$.

$OA_{\parallel}$ thus provides the best match to experiment; however, this configuration is *not* the lowest-energy structure, $AB_{N,\perp}$ being lower in energy by 0.19 eV. While this may seem like a small number, equilibrium occupation of the configurations at the temperatures of interest would result in the lower-energy configuration being overwhelmingly favored. We also expect sufficient accuracy (error bar smaller than ±0.1 eV) in our first-principles determination of this energy difference, because the relevant configurations differ only in the particular arrangement of the atomic

TABLE I. Calculated energy differences ΔE and lattice relaxations for Mg-H complexes in wurtzite GaN, for various lattice locations. The results were obtained from 96-atom supercell calculations. The global minimum ($AB_{N,\perp}$) is chosen as the zero of energy. Δd_{Mg} denotes the displacement of the Mg atom from its nominal lattice site, and Δd_N denotes the displacement of the N atom to which H is bonded, expressed as a percentage of the bond length. $d_{N\text{-}H}$ denotes the N-H distance in Å.

Configuration	ΔE (eV)	Δd_{Mg}	Δd_N	$d_{N\text{-}H}$ (Å)
$AB_{N,\perp}$	0.00	1.7%	14.6%	1.031
$OA_{\parallel}$	0.19	18.0%	10.6%	1.030
$AB_{N,\parallel}$	0.28	1.9%	12.2%	1.031
$OA_{\perp}$	0.31	10.9%	10.4%	1.038

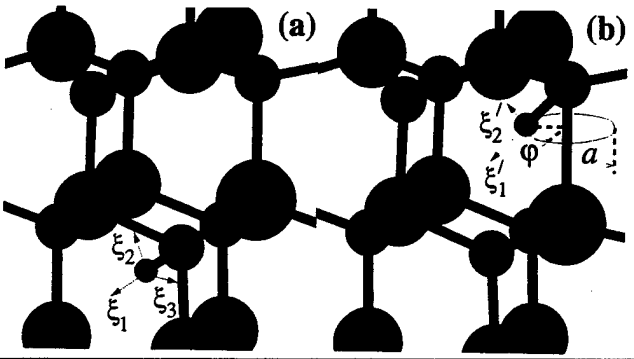


FIG. 2. Perspective view of Mg-H complexes in (a) the $AB_{N,\perp}$ and (b) the $OA_{\parallel}$ configuration. The path of H rotating around the crystal axis is also shown in (b). The directions of vibration for stretching (ξ_1) and bending modes (ξ_2, ξ_3) are indicated by arrows; in (b), ξ_3 is replaced by the azimuthal angle φ describing the rotational motion of H.

constituents. In addition, various checks confirmed the accuracy and reliability, including calculations using explicit inclusion of Ga $3d$ states instead of the nlcc. Interestingly, we obtained the same relative ordering of the energies in the *zinc-blende* structure, indicating that the overall symmetry of the system (hexagonal vs cubic) does not affect this result.

Next we investigated whether extrinsic effects might affect the relative stability of $AB_{N,\perp}$ and $OA_{\parallel}$. The experiments of Ref. [4] were performed on films grown on sapphire, and it is well known that such layers exhibit compressive stress due to the thermal mismatch between GaN and the substrate. We therefore calculated the energy difference between $AB_{N,\perp}$ and $OA_{\parallel}$ configurations when a 1% biaxial compression is applied, accompanied by an expansion of the lattice along the c axis. The magnitude of this strain is far larger than the strains typically present in epitaxial layers, which are closer to 0.1%. We found that the presence of this biaxial strain did not affect the energy difference between $AB_{N,\perp}$ and $OA_{\parallel}$.

The $OA_{\parallel}$ configuration actually does exhibit a characteristic that distinguishes it from $AB_{N,\perp}$: a low-energy excitation that qualitatively differs from the stretching and bending modes associated with the N-H bond. Our investigations indicate that it costs very little energy to move H away from the $OA_{\parallel}$ position, maintaining the "polar" angle θ with the c axis and changing the azimuthal angle φ [see Fig. 2(b)], effectively rotating the H atom around the c axis. If we allow complete relaxation, then for every H position the Mg and N atoms (as well as other host atoms) would assume different positions. The nature of the H motion actually makes it more likely that the surrounding (heavier) atoms will remain fixed, setting up a potential energy surface in which only the much lighter H atom moves. Either way, we find that the barrier for hydrogen rotation is on the order of 0.03 eV.

The presence of such a low-frequency excitation raises the possibility of a large entropy contribution to the free energy; we therefore explicitly calculate the free energy of the system. Since we require only the *difference* between the free energies of two distinct configurations of the Mg-H complex, we focus on those vibrational modes that are expected to be distinctly different in the two configurations, i.e., the modes directly associated with the H atom. For the $AB_{N,\perp}$ configuration, we need to consider one stretching and two bending modes [see Fig. 2(a)]. The calculated frequencies are 3017, 1061, and 1081 cm^{-1}. We express the free energy F in terms of the partition function Z: $F = -kT \ln Z$, with

$$Z = \prod_{i=1}^{3} \sum_{n_i=0}^{\infty} \exp\left(-\frac{E_{n_i}}{kT}\right), \tag{1}$$

where E_{n_i} are the energy eigenvalues for the stretching ($i = 1$) and bending ($i = 2,3$) mode oscillators. Typically, the harmonic approximation is used, in which $E_{n_i} = \hbar\omega_i(n_i + \frac{1}{2})$ and the sums in Eq. (1) can be evaluated analytically. However, since anharmonic contributions are important (evaluated as described in Ref. [10]), we use the full numerical expression. The three terms in Eq. (1) give rise to free-energy contributions F_1, F_2, and F_3, which are shown in Fig. 3(a). The finite values at $T = 0$ reflect zero-point energies. The decrease in F_i at higher T is due to entropy, which is larger for lower-frequency modes.

We then turn to evaluating the free energy for $OA_{\parallel}$. The stretching mode (ξ_1', at 3034 cm^{-1}) and one of the bending modes (ξ_2', at 1162 cm^{-1}) are very similar to those in $AB_{N,\perp}$ [see Fig. 2(b)]. The third excitation, however, is very different in nature, and corresponds to hydrogen rotating around the c axis, described by the azimuthal angle

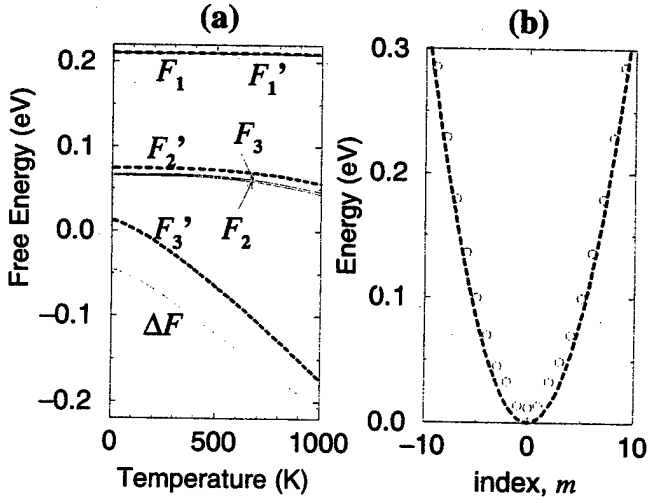


FIG. 3. (a) Free energies as a function of temperature. The thin solid lines indicate the free energy associated with the vibrational stretching (F_1) and bending ($F_{2,3}$) modes of Mg-H complexes in the $AB_{N,\perp}$ configuration; the dashed lines apply to the $OA_{\parallel}$ configuration, where F_3 corresponds to the rotational mode. The free-energy difference ΔF is shown as the thick solid line; it is almost entirely due to the difference between F_3 and F_3'. (b) Energy levels for the nearly free rotor in the $OA_{\parallel}$ configuration of the Mg-H complex, calculated using an energy barrier $E_{b\varphi} = 32$ meV and a radius $a = 0.79$ Å. The dashed line is a perfect parabola, reflecting the solutions for a free rotor.

φ. The potential energy in which the hydrogen moves shows some corrugation. The barrier for hydrogen rotation, when constraining the neighboring Mg and N atoms on the c axis, is calculated by determining the structure of the complex for an intermediate position of the hydrogen (corresponding to $\varphi = 30°$), and then fixing the Mg and host atoms in these positions and evaluating the energy as a function of the hydrogen coordinate. This yields a barrier of 32 meV. This value is not very sensitive to the detailed positions of the Mg and N atoms, which allow us to avoid the issue of coupling between the rotational motion and the other degrees of freedom.

Because the H atom sees a weak corrugation potential associated with the azimuthal angle φ, we can treat it dynamically as a nearly free rotor. The threefold symmetry around the c axis leads to $V_\varphi(\varphi + 120°) = V_\varphi(\varphi)$, allowing us to approximate the potential energy as $V_\varphi(\varphi) = E_{b\varphi}(\cos 3\varphi + 1)/2$, where $E_{b\varphi}$ is the calculated barrier. The φ-dependent part of the Schrödinger equation can be written as

$$HY(\varphi) = \left[-\frac{h^2}{2m_H a^2}\frac{\partial^2}{\partial\varphi^2} + V_\varphi(\varphi)\right]Y(\varphi) = EY(\varphi), \tag{2}$$

where a is the radius of the circular orbit [see Fig. 2(b)]. We have solved this equation using a plane-wave basis set [13]:

$$Y(\varphi) = \sum_{m=-\infty}^{\infty} c_m \frac{\exp(im\varphi)}{\sqrt{2\pi}}. \tag{3}$$

Including 51 plane waves ($m = -25$ to $m = 25$, corresponding to energy levels up to about 1 eV) suffices to converge the partition function. After diagonalizing the Hamiltonian matrix H we obtain the energy levels shown in Fig. 3(b). Note that the eigenvalues are very close to the parabola that would be obtained in the case of a free rotor, where $E_m = (h^2/2m_H a^2)m^2 = (0.0033\ \text{eV})m^2$.

Figure 3(a) shows that the free energy corresponding to the nearly free rotor, F_3', exhibits a large temperature dependence, reflecting the large entropy associated with this low-energy excitation. Figure 3(a) also shows the free-energy difference $\Delta F = F' - F$ between the $OA_\parallel$ and the $AB_{N,\perp}$ configurations; F_3' obviously makes the largest contribution to ΔF. We see that ΔF equals -0.19 eV (the energy difference between $OA_\parallel$ and $AB_{N,\perp}$) at around $T = 900$ K or 627 °C. At higher temperatures, the $OA_\parallel$ configuration will be stabilized. In the experiments of Ref. [4] the hydrogen was introduced during metal-organic chemical vapor deposition at 1000 °C, a temperature high enough to favor the $OA_\parallel$ configuration.

An intriguing consequence of our results is that, given the small value of the barrier $E_{b\varphi}$, the hydrogen will tend to be quantum-mechanically delocalized. Indeed, in the limit of a free rotor ($E_{b\varphi} = 0$), hydrogen would exhibit equal probability to occur at any value of the azimuthal coordinate φ even at $T = 0$. Whether or not this delocalization as a function of φ occurs at the low temperatures (5 K) of the experiment in Ref. [4] is a question we feel cannot be answered without a full quantum-mechanical treatment of the hydrogen motion. However, we have checked that the procedure used in Ref. [4] to arrive at the bond angle $\theta = 130°$ is equally valid even if H has equal probability of occurring at any angle φ. The presence of rotational excitations with an energy on the order of a few meV could lead to broadening of the observed vibrational line (1 meV $\approx$ 8 cm^{-1}). Substantial broadening was indeed observed in the spectra reported in Ref. [4].

In summary, we have presented a model for the structure of the Mg-H complex in GaN that agrees with the unusual bond angle that has been experimentally observed [4]. The structure exhibits a set of low-energy excitations, consisting of a nearly free rotor with a very small barrier for rotation. The entropy associated with this motion plays an essential role in stabilizing the complex. Our work highlights the importance of taking finite-temperature free-energy effects into account when assessing the relative stability of various possible configurations of a defect, impurity, or complex.

This work was supported by the Air Force Office of Scientific Research, Contract No. F4920-00-C-0019, monitored by G. Witt. We thank N. M. Johnson, J. Neugebauer, and M. D. McCluskey for useful discussions and suggestions.

*Permanent address: School of Physics, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand.

[1] *Hydrogen in Semiconductors,* edited by J. I. Pankove and N. M. Johnson, Semiconductors and Semimetals Vol. 34 (Academic Press, San Diego, 1991).
[2] Ref. [1], Chap. 13.
[3] W. Götz *et al.,* Appl. Phys. Lett. **69**, 3725 (1996).
[4] B. Clerjaud *et al.,* Phys. Rev. B **61**, 8238 (2000).
[5] J. Neugebauer and C. G. Van de Walle, Phys. Rev. Lett. **75**, 4452 (1995).
[6] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964); W. Kohn and L. J. Sham, *ibid.* **140**, A1133 (1965).
[7] M. Bockstedte, A. Kley, J. Neugebauer, and M. Scheffler, Comput. Phys. Commun. **107**, 187 (1997).
[8] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).
[9] S. Limpijumnong and C. G. Van de Walle, in *GaN and Related Alloys,* edited by U. Mishra *et al.,* MRS Symposia Proceedings No. 639 (Materials Research Society, Pittsburgh, 2001), p. G4.3.
[10] C. G. Van de Walle, Phys. Rev. Lett. **80**, 2177 (1998).
[11] S. Limpijumnong and C. G. Van de Walle (unpublished).
[12] S. M. Myers *et al.,* J. Appl. Phys. **88**, 4676 (2000).
[13] E. Artacho *et al.,* Phys. Rev. B **56**, 3820 (1997).

First-Principles Surface Phase Diagram for Hydrogen on GaN Surfaces

Chris G. Van de Walle[1,2,3] and J. Neugebauer[2]

[1]*Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, California 94304*
[2]*Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin-Dahlem, Germany*
[3]*Paul-Drude-Institut, Hausvogteiplatz 5-7, D-10117 Berlin, Germany*

(Received 27 August 2001; published 28 January 2002)

We discuss the derivation and interpretation of a generalized surface phase diagram, based on first-principles density-functional calculations. Applying the approach to hydrogenated GaN surfaces, we find that the Gibbs free energies of relevant reconstructions strongly depend on temperature and pressure. Choosing chemical potentials as variables results in a phase diagram that provides immediate insight into the relative stability of different structures. A comparison with recent experiments illustrates the power of the approach for interpreting and predicting energetic and structural properties of surfaces under realistic growth conditions.

DOI: 10.1103/PhysRevLett.88.066103 PACS numbers: 68.35.–p, 68.43.–h, 81.10.–h

A detailed understanding of semiconductor surface reconstructions is essential for controlling growth and materials properties. In the case of GaN, where materials quality is still limiting device progress, the investigations until now have focused on reconstructions of *bare* GaN surfaces [1]. The present study focuses on the role of hydrogen (H), which is important because H is abundantly present in the most commonly used growth techniques for nitride semiconductors, including metal-organic chemical vapor deposition (MOCVD), hydride vapor-phase epitaxy (HVPE), and molecular-beam epitaxy (MBE) when NH_3 is used as the nitrogen source.

We have therefore performed detailed investigations of H interactions with GaN surfaces based on state-of-the-art pseudopotential-density-functional calculations. We focus on the technologically most relevant GaN(0001) surface, which is the polarity observed during MOCVD of GaN on sapphire as well as MBE on Si-face SiC. Since we are particularly interested in consequences for growth, it is essential to take finite temperature effects into account, requiring the calculation of free energies. This distinguishes our approach from previous work in which only zero-temperature energies were calculated, and then only for a small number of structures [2–4]. We will show that the energetic and structural features of the surface reconstructions dramatically depend on temperature (T) and pressure (p).

In thermochemistry, the Gibbs free energy G is usually expressed as a function of T and partial pressures. An alternative approach is to use *chemical potentials* μ as variables and to express G as a function of p, T, and all independent μ's. For the system at hand, the Gibbs' Phase Rule produces four degrees of freedom, resulting in a four-dimensional phase space that is difficult to analyze or visualize. Our detailed analysis will show that the *explicit* dependence on p and T is small and can be neglected. The surface energy can then be expressed solely as a function of the chemical potentials μ_{Ga} and μ_{H}, while the p and T dependence is implicit. The advantage of this methodology is its physically intuitive character. Comparison with experiment requires only that the chemical potentials be evaluated as a function of T and partial pressures. The power of this approach will become evident when we use it to interpret the recent *in situ* observations of surface reconstructions based on grazing incidence x-ray scattering by Munkholm *et al.* [5]. We will show that their measured transition between two reconstructions coincides with a specific boundary in our phase diagram, offering direct information about the nature of the observed structures.

Our first-principles approach is based on density-functional theory in the local-density approximation (LDA) [6], using *ab initio* pseudopotentials [7]. Test calculations using the generalized gradient approximation produced results in close agreement with LDA. The Ga $3d$ states are explicitly included as valence states, requiring a 70 Ry plane-wave cutoff. We use a slab geometry, consisting of at least four double layers of GaN plus at least 7 Å of vacuum. The lower surface of the slab is passivated with fractionally charged H, and the positions of all atoms in the lower half of the slab are kept fixed. Convergence tests indicated that these choices allow us to determine energy differences with an accuracy of better than 0.05 eV.

We choose the ideal, bare 1×1 surface (relaxed but unreconstructed) as our reference, and define the *Gibbs free energy of formation* ΔG^f as follows:

$$\Delta G^f = E_{\mathrm{tot}}[\mathrm{GaN(0001)}] - E_{\mathrm{tot}}[\mathrm{GaN(0001), ideal}] + \Delta F_{\mathrm{vib}} - n_{\mathrm{Ga}}\mu_{\mathrm{Ga}} - n_{\mathrm{N}}\mu_{\mathrm{N}} - n_{\mathrm{H}}\mu_{\mathrm{H}}. \quad (1)$$

$E_t[\mathrm{GaN(0001)}]$ is the calculated total energy for the surface under study, and $E_{\mathrm{tot}}[\mathrm{GaN(0001), ideal}]$ is the total energy of our reference system. $n_{\mathrm{Ga(N,H)}}$ is the number of Ga(N, H) atoms added to the unit cell (positive or negative). ΔF_{vib} includes vibrational contributions to the free energy and is discussed below. μ_{Ga} and μ_{N}

 © 2002 The American Physical Society

are the chemical potentials of Ga and N, i.e., the free energies of the reservoirs with which Ga and N atoms are exchanged. Invoking equilibrium with GaN [8] leaves us with a single parameter to describe the stoichiometry, for which we choose μ_{Ga}. μ_{Ga} varies over the thermodynamically allowed range: $\mu_{Ga[bulk]} + \Delta H_f[GaN] < \mu_{Ga} < \mu_{Ga[bulk]}$, the upper limit corresponding to Ga-rich conditions, the lower limit to N-rich ($\mu_N = \mu_{N[N_2]}$). $\Delta H_f[GaN]$ is the enthalpy of formation (negative for a stable compound). Our calculated value for $\Delta H_f[GaN]$ is -1.24 eV (experiment: -1.17 eV, Ref. [9]). μ_H, finally, is the free energy of H_2 and describes the abundance of H in the environment.

Figure 1(a) shows our calculated formation energies as a function of μ_{Ga} at $T = 0$. In total, we examined over 30 different structures; Fig. 1 includes only those reconstructions (schematically illustrated in Fig. 2) that were found to be energetically favorable in some part of the phase space spanned by the chemical potentials. In the absence of H, we reproduce the results for bare GaN(0001) surfaces reported in Ref. [1]; specifically, under moderately Ga-rich conditions the most favorable reconstruction is the Ga adatom in a $T4$ position (i.e., above a second-layer N atom), while under N-rich conditions the N adatom in the $H3$ position (i.e., the "hollow site" above the hexagonal channel) is most stable. Under extreme Ga-rich conditions, a Ga-bilayer reconstruction is energetically most favorable [10].

Although we made no *a priori* assumptions, we find *a posteriori* that the lowest-energy structures all obey *electron counting* [11]: any N dangling bonds (dbs) that remain on the surface after formation of N-H or Ga-H bonds are filled with two electrons, while Ga dangling bonds are empty. We actually examined many structures with 1×1 and $\sqrt{3} \times \sqrt{3}$ periodicity, in addition to the 2×2 structures reported in Fig. 1. Only the 2×2 structures were energetically favorable, consistent with the easily derived result that the electron counting rule can be satisfied for tetrahedral coordination only if the unit cell contains a multiple of four surface atoms.

Figure 1(a) shows that at $T = 0$ hydrogenated surfaces with a large number of N-H bonds have significantly lower energies than the bare surfaces. This is consistent with the notion that N-H bond formation is favorable due to the large bond strength. In particular, the $NH_3 + 3NH_2$ structure is favored over almost the entire range of μ_{Ga}; this structure is formed by adding one NH_3 and three NH_2 molecules in on-top positions (see Fig. 2). Alternatively, one can think of this structure as a N-terminated GaN(0001) structure, with 9 of the 12 N dbs passivated by H; this number of H atoms allows for the electron counting rule to be satisfied. Under extreme Ga-rich conditions, the NH_3 + 3Ga-H structure is favorable; here, one NH_3 unit sits on top of one of the Ga surface atoms, and Ga-H bonds are formed at the other three Ga atoms in the cell. This structure is very close in energy to a stoichiometrically equivalent structure, namely, V_{Ga} + 6H, which consists of a Ga vacancy in which all three N dbs are saturated with H, plus three Ga-H bonds.

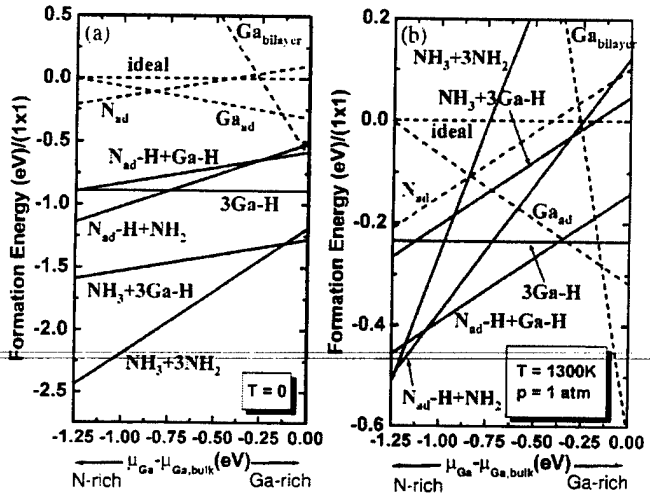


FIG. 1. Calculated formation energies for GaN(0001) surfaces, as defined in Eq. (1), as a function of μ_{Ga}. All structures (except $Ga_{bilayer}$) have 2×2 periodicity, but energies are expressed in eV per 1×1 cell. Dashed lines indicate hydrogen-free surfaces. (a) $T = 0$; (b) $T = 1300$ K, $p(H_2) = 1$ atm. Note the very different energy scale in the two plots.

If we limited our investigation to the $T = 0$ results depicted in Fig. 1(a), we would conclude that the NH_3 + $3NH_2$ structure, which includes the largest number of N-H bonds, is dominant on the hydrogenated GaN(0001) surface. We will see, however, that this conclusion is not valid at higher temperatures if finite-temperature effects are incorporated in the formation energy defined in Eq. (1). The temperature dependence arises from vibrational contributions to the energy and entropy, and can, in principle, be evaluated entirely from first principles based on a calculation of the vibrational spectrum. Inspection of the vibrational modes calculated for a subset of structures reveals, however, that ΔF_{vib} in Eq. (1) can be calculated by assuming specific frequencies for the vibrational modes of the bonds that participate in these surface reconstructions [12]. Assuming that vibrational frequencies do not change by more than ± 100 cm^{-1} due to changes in the environment, we obtain a conservative error bar of less than 0.1 eV

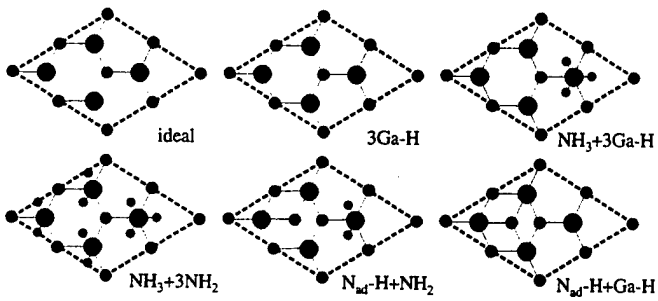


FIG. 2. Schematic top view of prevalent 2×2 reconstructions for GaN(0001) surfaces. Large open circles represent Ga atoms, solid circles N, and small open circles H.

on the free energies calculated using this approximation. The temperature (and pressure) dependence of the reservoirs has also been included in Eq. (1); this is particularly important for gaseous species, e.g., for hydrogen:

$$2\mu_H = E_{H_2} + kT\left[\ln\left(\frac{pV_Q}{kT}\right) - \ln Z_{rot} - \ln Z_{vib}\right], \quad (2)$$

where E_{H_2} is the energy of an H_2 molecule, k is the Boltzmann constant, T is the temperature, and p is the pressure. $V_Q = (h^2/2\pi mkT)^{3/2}$ is the quantum volume, and Z_{rot} and Z_{vib} are the rotational and vibrational partition functions. The temperature dependence of μ_H for two different pressures is shown in Fig. 3(b).

Figure 1(b) shows the formation energies calculated at $T = 1300$ K (a typical MOCVD growth temperature) and $p(H_2) = 1$ atm. Not only are the energetically favored reconstructions entirely different from those at $T = 0$, but the energy difference between hydrogenated and bare surfaces has been dramatically reduced. While at a low temperature several eV could be gained by forming N-H bonds, at a high temperature the energy difference between hydrogenated and bare surfaces has been reduced to less than 0.3 eV—with the H-free Ga_{ad} and $Ga_{bilayer}$ structures actually being favored under Ga-rich conditions. The main reason for this dramatic difference is the large change in the value of μ_H, i.e., the free energy of the reservoir with which H atoms are exchanged: at 1300 K μ_H is about 1 eV lower than at $T = 0$ [Eq. (2) and Fig. 3(b)]. A lower value of μ_H means that less energy can be gained by taking a H atom out of the reservoir and binding it to the GaN surface. The vibrational contributions to the free energy included in ΔF_{vib} are significantly smaller, on the order of 0.1 eV at $T = 1300$ K.

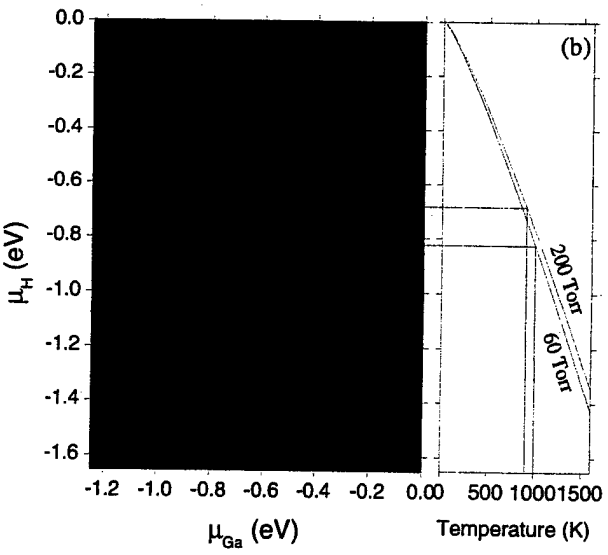


FIG. 3. (a) Phase diagram for the GaN(0001) surface in the presence of H, as a function of μ_{Ga} and μ_H. $\mu_H = 0$ corresponds to H_2 molecules at $T = 0$; $\mu_{Ga} = 0$ corresponds to bulk Ga. Dots indicate experimental data from Ref. [5]; within the error bars, these data agree with the calculated NH_3 + 3Ga-H/3Ga-H phase boundary highlighted by the thicker line. Note that the V_{Ga} + 6H structure is stoichiometrically and energetically equivalent to NH_3 + 3Ga-H. (b) Temperature dependence of μ_H for two different pressures [Eq. (2)].

Changes in T (and p) clearly have dramatic effects on the relative stability of different structures. As mentioned in the introduction, a traditional phase diagram would reflect four degrees of freedom, based on $\Delta G^f(p, T, p_{Ga}, p_H)$. The four-dimensional nature of the diagram presents a serious obstacle to straightfoward interpretation of the first-principles results. We now show that the number of degrees of freedom can be reduced to *two* if we choose the *chemical potentials* as variables, as shown in Fig. 3(a). The diagram shown here was calculated at $T = 950$ K (in order to facilitate the comparison with experiment discussed below), but the key point is that the important qualitative features of the diagram are *independent of temperature and pressure*. Small quantitative shifts in the positions of the boundaries do occur, but these amount to only about 0.1 eV along either μ_H or μ_{Ga}. Obviously, to obtain rigorous numbers one can refer back to Eq. (1) and take the detailed T and p dependence into account—but important qualitative insights can reliably be obtained from Fig. 3(a), even at temperatures significantly different from $T = 950$ K.

The reasons why the number of degrees of freedom can be reduced from four to two are as follows. First, for the system at hand the *total pressure* plays a negligible role. As already implied in Eq. (1), the $p\Delta V$ term can be safely neglected in the free energy for the surface. In addition, the chemical potentials depend only on the respective *partial* pressures. Dropping the *temperature* as an explicit degree of freedom is based on the more approximate argument that the vibrational free energy contributions to ΔG are small, and exhibit significant cancellation when comparing ΔG for different reconstructions; the net effect of temperature on the phase boundaries is therefore quite small. We can thus write $\Delta G^f(p, T, p_{Ga}, p_H) \approx \Delta G^f[\mu_{Ga}(p_{Ga}, T), \mu_H(p_H, T)]$. The remaining two variables, μ_{Ga} and μ_H, are of course functions of T and p, but otherwise the phase diagram exhibits no explicit T or p dependence. We suggest that this approach is quite general in nature. Hydrogenated GaN is actually a fairly extreme case, with both very high frequencies (N-H bonds, contributing to sizable vibrational energies) and very low frequencies (torsion modes, contributing to sizeable vibrational entropies). The success of the approach for this extreme case indicates it could be usefully applied to other systems as well.

We now discuss how Fig. 3 can be used to relate surface structures to experimental growth conditions. At the bottom of the diagram (low μ_H) we find the structures that dominate when little or no H is present. At the top of the diagram ($\mu_H = 0$), we find the structures that were stable at $T = 0$ [see Fig. 1(a)]. A horizontal line at $\mu_H = -0.98$ eV would cut through the structures that were found to be stable at 1300 K and $p(H_2) = 1$ atm [Fig. 1(b)]. The temperature dependence of μ_H is shown

in Fig. 3(b), for two different pressures. Figure 3 makes it straightforward to investigate which structures would be stabilized by raising or lowering the pressure or the temperature.

Munkholm *et al.* [5], performed *in situ* grazing incidence x-ray scattering on GaN(0001) surfaces in an MOCVD environment. Starting from 2-μm-thick GaN films grown on sapphire, they grew a 30-nm-thick layer at 1000 °C, on which they then performed their search for surface reconstructions. During this search, the sample was held at temperatures up to 1000 °C, and $p(NH_3)$ was varied between 2 and 140 Torr. The flow rate of the NH_3/carrier gas mixture was kept constant, with a total pressure of 200 Torr. Results obtained with N_2 as the carrier gas were very similar to those obtained for H_2.

At high temperatures, only a 1×1 reconstruction was observed. At lower temperatures and sufficiently low NH_3 pressures, a different reconstruction was seen, which was identified as $(\sqrt{3} \times 2\sqrt{3})R30°$. The transition between the two reconstructions was mapped out as a function of T and $p(NH_3)$. From the temperature dependence of $p(NH_3)$ at the transition, an apparent activation energy of 3.0 ± 0.2 eV was extracted. We now show that this experimentally established activation energy can be directly correlated with the energy difference between two surface reconstructions.

The experimental observation that the phase boundary is insensitive to whether H_2 or N_2 is used as the carrier gas indicates that the transformation between the two observed phases occurs purely through equilibration with NH_3. The equilibrium expression $\mu_N + 3\mu_H = \mu_{NH_3}$ then immediately fixes the *slope* of the phase boundary, already severely restricting the reconstructions that could be involved in the experimental transition. The position of the boundary is further determined by using the actual value $\mu_{NH_3} = -2.28 \pm 0.2$ eV along the experimental phase boundary (referenced to the energy of the NH_3 molecule), and has been included in Fig. 3(a). The end points correspond to the temperatures and pressures at the limits of the range explored in Ref. [5], assuming H_2 as the carrier gas. Within the experimental [5] and computational uncertainty, we find good agreement with the calculated NH_3 + 3Ga-H/3Ga-H phase boundary, including the fact that at temperatures above those reported in Ref. [5] this particular phase boundary should disappear because another reconstruction (N_{ad}-H + Ga-H) becomes more favorable.

The experimental observations are therefore consistent with a transition between two structures that differ only in the addition/subtraction of one NH_3 unit in the surface unit cell. The 3Ga-H structure (corresponding to a 3/4 monolayer coverage of H) is expected to exhibit 1×1 symmetry, since the experiment is not sensitive to H atoms [5]; in addition, disorder (due to different H positions in each 2×2 cell) would reduce the symmetry to 1×1. As for the other reconstruction, we suggest it is unlikely that it would indeed exhibit a $(\sqrt{3} \times 2\sqrt{3})R30°$ symmetry, as proposed in Ref. [5], since such a unit cell would not contain a multiple of four atoms and therefore would be energetically highly unfavorable, according to our first-principles results. The reconstruction peaks observed in Ref. [5] in fact exhibited $(2\sqrt{3} \times 2\sqrt{3})R30°$ symmetry, and the reduction to $(\sqrt{3} \times 2\sqrt{3})R30°$ was based on a fit. A $(2\sqrt{3} \times 2\sqrt{3})R30°$ reconstruction could be constructed out of 2×2 *motifs,* with interactions between the adspecies in neighboring 2×2 units stabilizing the structure. Explicit investigations of structures with 12 atoms in the surface unit cell are beyond the scope of the present investigations, but we suggest that the NH_3 + 3Ga-H structure (and/or the V_{Ga} + 6H structure, which is stoichiometrically and energetically equivalent) is a basic building block of the larger reconstruction.

In summary, we have outlined a computational approach for obtaining phase diagrams for systems consisting of an impurity-covered surface in thermodynamic contact with a gaseous environment, as characteristic of MOCVD or HVPE. The approach gives immediate access to the atomic structure and phase transitions of surfaces under realistic growth conditions, as illustrated with the important example of hydrogenated GaN surfaces. The methodology can be easily generalized to surface alloying or oxidation.

This work was supported in part by AFOSR (Contract No. F4920-00-C-0019), by the Alexander von Humboldt Foundation, and by the Deutsche Forschungsgemeinschaft (Schwerpunktprojekt "Group-III Nitrides"). We are grateful to J. Northrup, R. Feenstra, M. Scheffler, and O. Brandt for useful discussions and to A. da Rosa for assistance with the calculations.

[1] A. R. Smith *et al.*, Phys. Rev. Lett. **79**, 3934 (1997).
[2] J. Fritsch, O. F. Sankey, K. E. Schmidt, and J. B. Page, Phys. Rev. B **57**, 15 360 (1998).
[3] K. Rapcewicz, M. Buongiorno Nardelli, and J. Bernholc, Phys. Rev. B **56**, R12 725 (1997).
[4] J. Elsner, M. Haugk, G. Jungnickel, and Th. Frauenheim, Solid State Commun. **106**, 739 (1998).
[5] A. Munkholm *et al.*, Phys. Rev. Lett. **83**, 741 (1999).
[6] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964); W. Kohn and L. J. Sham *ibid.* **140**, A1133 (1965).
[7] M. Bockstedte, A. Kley, J. Neugebauer, and M. Scheffler, Comput. Phys. Commun. **107**, 187 (1997).
[8] MOCVD growth conditions are close enough to equilibrium (as established by sufficiently high mobilities of point defects on the surface and in the bulk) for this assumption to be valid.
[9] *CRC Handbook of Chemistry and Physics,* edited by David R. Lide (CRC Press, Boca Raton, Florida, 1992), 73rd ed., pp. 5–18.
[10] J. E. Northrup, J. Neugebauer, R. M. Feenstra, and A. R. Smith, Phys. Rev. B **61**, 9932 (2000).
[11] M. D. Pashley, Phys. Rev. B **40**, 10 481 (1989).
[12] J. E. Northrup, R. Di Felice, and J. Neugebauer, Phys. Rev. B **56**, R4325 (1997).

phys. stat. sol. (b) **228**, No. 1, 303–307 (2001)

Passivation and Doping due to Hydrogen in III-Nitrides

S. LIMPIJUMNONG[1]) *) and C. G. VAN DE WALLE

Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, USA

(Received June 29, 2001; accepted August 4, 2001)

Subject classification: 61.72.–y; 71.55.Eq; S7.14

We have systematically studied the electronic structure and stability of hydrogen in AlN, GaN, and InN, based on first-principles calculations. In GaN and AlN, H is amphoteric and always compensates the prevailing conductivity: in GaN, H^+ is stable for Fermi levels below 2.2 eV, and in AlN, H^+ is stable for E_F below 2.5 eV. In InN, we find that H^+ is stable for *all* Fermi level positions; i.e., H behaves exclusively as a donor. Consequences for controlling the conductivity of InN are discussed.

Introduction Hydrogen is a common impurity in III-nitrides since it is abundantly present in the growth environment of many techniques used to grow nitrides. It is now well known that H passivates acceptors during growth of p-type GaN, necessitating a post-growth anneal to activate the acceptors, for a review, see [1]. The behavior of H in GaN has previously been addressed with first-principles calculations, producing detailed information about the stability and microscopic structure of different charge states [2]. Device applications of nitrides, however, almost always involve alloying with AlN and/or InN. Knowledge of the behavior of H in these other compounds is therefore essential. In this work, we systematically study the behavior of H in the three nitride compounds. The behavior in AlN turns out to be qualitatively similar to GaN: H is amphoteric and always compensates the prevailing conductivity. Hydrogen thus behaves as a donor in p-type GaN or AlN, and as an acceptor in n-type material. InN, however, exhibits a totally different behavior, with H acting exclusively as a donor.

The tendency for H to act as a donor in p-type material forms the driving force for *passivation* of acceptors. Vibrational spectroscopy of acceptor–hydrogen complexes provides a powerful way of monitoring the presence of H and the degree of acceptor activation [3]. A comparison with first-principles calculations then allows for an unambiguous identification of the nature and microscopic structure of the complex. We have recently performed a comprehensive investigation of structures and vibrational properties of acceptor–hydrogen complexes in GaN, taking proper account of anharmonic effects (which are large). The results, including a novel configuration for the Mg–H complex, are discussed elsewhere [4]. In this paper, we focus on the behavior of H as an isolated interstitial impurity.

Methods Our calculations are based on density-functional theory within the local density approximation, using ab initio pseudopotentials with a plane-wave basis set [5]. For GaN and InN we use the so-called "non-linear core correction" (nlcc) [6], with an energy cutoff of 40 Ry. Explicit inclusion of the Ga 3d, and In 4d electrons leads to

[1]) Corresponding author; e-mail: SXL57@po.cwru.edu
*) Permanent address: School of Physics, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand.

© WILEY-VCH Verlag Berlin GmbH, 13086 Berlin, 2001 0370-1972/01/22811-0303 $ 17.50+.50/0

very similar results for the configurations studied here, as reported in Ref. [7]. In this study we focus on the zinc-blende phase of the III-nitrides; results for the wurtzite phase (which differs in local environment only beyond third nearest neighbors) are expected to be very similar. Our calculations are performed at the theoretical lattice constants: 4.30 Å for AlN, 4.38 Å for GaN, and 5.10 Å for InN. These are in satisfactory agreement with the experimental values of 4.40, 4.50, and 5.00 Å [8]. Our calculated band gaps are 3.03 eV for GaN, −0.23 eV for InN, and 4.60 eV (direct) and 3.26 eV (indirect) for AlN. As usual for density-functional calculations, these values underestimate the actual gaps: 3.2 eV for GaN [8], 1.8 eV for InN (based on the calculated difference between the zinc-blende and wurtzite gaps [9]), and 5.0 eV for AlN (indirect gap from quasiparticle calculations [10]). The band-gap problem should not affect our conclusions, provided care is taken in the interpretation of the results, as discussed below. For the impurity calculations we used supercells containing 32 and 64 atoms, with at least 24 host atoms relaxed. Brillouin-zone integration was carried out with a $2 \times 2 \times 2$ mesh in the reciprocal unit cell, reduced by symmetry to a set of two irreducible **k** points.

The *formation energy* E^f determines the concentration c of an impurity through the expression: $c = N_{\text{sites}} \exp(-E^f/kT)$. N_{sites} is the number of sites in the lattice (per unit volume) on which the impurity can be incorporated, k is the Boltzmann constant, and T is the temperature. The formation energy of an interstitial H atom in charge state q is defined as

$$E^f(H^q) = E_{\text{tot}}(H^q) - E_{\text{tot}}(\text{bulk}) - \mu_H + qE_F, \tag{1}$$

where $E_{\text{tot}}(H^q)$ is the total energy of H^q at a specific interstitial site, E_{tot} (bulk) is the total energy of a bulk supercell of the same dimensions as the one used to perform the impurity calculations, and μ_H is the energy of the hydrogen reservoir, i.e., the H chemical potential. For purposes of displaying our results we choose μ_H to be the energy of an H_2 molecule at $T = 0$. No zero-point energies are included. E_F is the Fermi level. For more information and calculational details we refer to Refs. [5] and [11].

Hydrogen in GaN Our calculations here duplicate those published in Ref. [2]; they were carried out to ensure consistency when comparing with AlN and InN. The comprehensive exploration of the complete energy surface in the earlier work highlighted three interstitial locations where H can be stable or metastable: the bond-center site (BC), the antibonding site near a nitrogen atom (AB_N), or the antibonding site near a gallium atom (AB_{Ga}).

For *positively* charged ($\underline{H}^+$) we find almost equal formation energies at BC and AB_N sites, AB_N being marginally lower. The N–H bond lengths are 1.03 Å for BC and

Table 1
Formations energies (in eV) (at $E_F = 0$) and stable for H in nitride semiconductors

charge state	stable site	AlN	GaN	InN
H^+	AB_N	−1.04	−0.30	−1.11
H^0	AB_{cation}	2.69	3.09	2.03
H^-	AB_{cation}	3.85	4.09	2.55

1.05 Å for AB_N. The AB_{Ga} site is 2.38 eV higher in energy. For *neutral* hydrogen ($\underline{H^0}$), AB_{Ga} and BC are comparable in energy, with AB_N only 0.69 eV higher. Note that the Ga–H bond length is such that the AB_{Ga} almost coincides with the tetrahedral interstitial (T_d) site. For *negatively* charged hydrogen ($\underline{H^-}$), the AB_{Ga} site is strongly preferred; the energy is higher by 2.21 eV at BC and by 3.83 eV at AB_N.

The stable sites and corresponding formation energies are tabulated in Table 1, and the formation energies as a function of E_F are shown in Fig. 1b. H^+ is energetically stable for E_F below $\varepsilon(+/-) = 2.20$ eV, while H^- is more stable for higher Fermi levels. This means that in p-type GaN H^+ is stable, while in n-type GaN H^- is stable, providing the basis for hydrogen's tendency to always act as a compensating center. H^0 is never stable, which is the characteristic of a "negative-U" system. The value of U is equal to $\varepsilon(0/-) - \varepsilon(+/0) = -2.39$ eV. The origin of this large negative U in GaN has been discussed in Ref. [2].

Hydrogen in AlN The lattice constant of AlN is slightly smaller than that of GaN, and the elastic constants slightly larger [8], causing the BC site to be less favorable. We find, in fact, that BC is no longer a local minimum for $\underline{H^+}$: a slight displacement of H off the BC site causes it to relax towards the AB_N site. AB_N is the energetically most stable site, with a N–H bond length of 1.06 Å. AB_{Al} is higher in energy by 2.40 eV; the latter difference is essentially equal to the value in GaN. For $\underline{H^0}$, AB_{Al} is most stable, with H sitting very close to the T_d site. AB_N is 0.40 eV higher in energy, and BC is 0.78 eV higher; again we see (just like in GaN) that the energy differences between these sites are relatively small in the case of H^0. Negatively charged hydrogen ($\underline{H^-}$), finally, prefers to stay close to the cations and strongly prefers AB_{Al}. The Al atoms move toward H, making the Al–H bond length 5% shorter than Al–N. The energy of H^- at BC and AB_N sites is much higher (by 3.42 and 3.50 eV) than at the AB_{Al} site. The formation energies as functions of Fermi level are shown in Fig. 1a.

Overall, the behavior of H in AlN is very similar to the case of GaN with $\varepsilon(+/-)$ now at 2.45 eV, and a slightly larger magnitude of $U = -2.57$ eV. Specifically, the general picture presented in Ref. [2] holds true for AlN, which has an even wider band gap

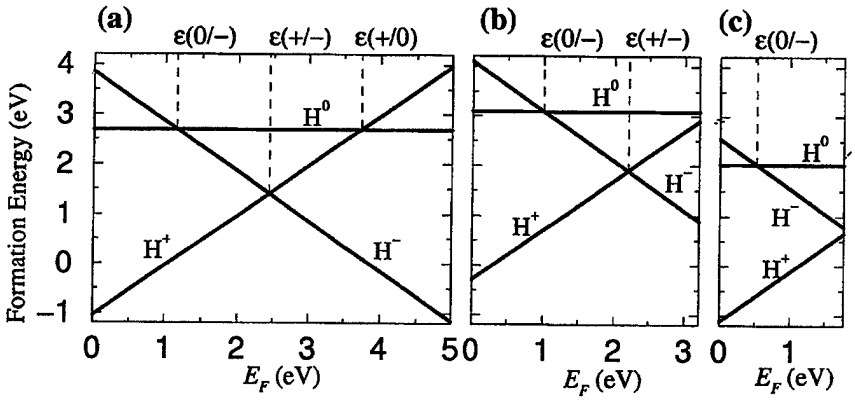


Fig. 1. Formation energies as a function of Fermi level for H^+, H^0, and H^- in a) AlN, b) GaN, and c) InN. $E_F = 0$ corresponds to the valence-band maximum, and formation energies are referenced to the energy of an H_2 molecule

than GaN. The preference for different sites in different charge states can be explained on the basis of Coulomb interactions: H^+ prefers regions of high charge density and thus stays close to the anion (N), favoring AB_N or BC. On the other hand, H^- prefers to stay away from the anion and closer to the cation, making AB_{Ga} or AB_{Al} the preferred site. Both H^+ and H^- exhibit large energy differences between sites close to anions versus cations, while the energy surface for H^0 is much "flatter".

Although the behavior of H in AlN is very similar to GaN, the larger gap of AlN leads to an interesting effect. In GaN, H^- is never quite as stable as H^+, due to an asymmetric placement of $\varepsilon(+/-)$ in the gap (see Fig. 1b). In AlN, however, the larger gap can lead to E_F positions far exceeding $\varepsilon(+/-)$, leading to a higher possible stability of H^-. We therefore predict that H will play a more important role in n-type AlN than in n-type GaN.

Hydrogen in InN The band gap of InN is much smaller than that of AlN and GaN (1.8 eV in the zinc-blende structure), and we will see that this leads to an interesting behavior. For $\underline{H^+}$, the AB_N site is favored again, and BC is higher in energy by only 0.18 eV. At AB_{In}, the total energy is higher by 2.95 eV. For $\underline{H^0}$ and $\underline{H^-}$, great care has to be taken in the interpretation of the density-functional calculations (as discussed previously for GaN in Ref. [7]). We carefully checked the nature of the energy level in the band structure that is occupied with one or two electrons in the H^0 and H^- calculations. For AB_{In}, the corresponding state is indeed localized near the H atom. For AB_N and BC, however, this state turns out to be an *extended* state. Correct calculations for H^0 and H^- should therefore *not* place electrons in this state – the actual hydrogen-induced energy level occurs higher in the band structure. When these considerations are taken into account, AB_N and BC are found to be higher in energy than AB_{In} for both H^0 and H^-. The formation energies as a function of Fermi level for all three charge states are shown in Fig. 1c. Again, U is large and negative ($U = -2.62$ eV), and the value of $\varepsilon(+/-)$ is 1.83 eV. Since the energy gap of InN is smaller than $\varepsilon(+/-)$, the positive charge state is the *only* stable charge state for H in InN, regardless of Fermi-level position. Note that band-gap corrections would *raise* the energy of H^- (the charge state with occupied gap levels) with respect to H^+, thus strengthening our conclusions about the stability of H^+. Therefore, hydrogen always acts as a *donor* in InN.

This behavior is surprising, since it differs from the usual amphoteric character of H in other semiconductors (including GaN and AlN). In InN, H can act as a *source* of n-type conductivity. This is important to keep in mind when growing InN; growth techniques such as metal-organic chemical vapor deposition and hydride vapor phase epitaxy introduce hydrogen from source gases and carrier gases, and even in molecular beam epitaxy H can be present either intentionally (when NH_3 is used as the N source) or unintentionally (as a hard-to-avoid background impurity). We have previously discussed other sources of n-type conductivity in InN [12]: nitrogen vacancies are too high in energy to be a concern, but unintentional impurities such as silicon or oxygen do act as donors and can be easily incorporated. Hydrogen should be added to this list of unintentional donors in InN.

Acknowledgements This work was supported by the Air Force Office of Scientific Research, Contract No. F4920-00-C-0019, monitored by G. Witt. We thank J. E. Northrup, N. Johnson, and J. Neugebauer for useful discussions and suggestions.

References

[1] C. G. VAN DE WALLE and N. M. JOHNSON, in: Gallium Nitride (GaN) II, Eds. J. I. PANKOVE and T. D. MOUSTAKAS, Semiconductors and Semimetals, Vol. 57, Treatise Eds. R. K. WILLARDSON and E. R. WEBER, Academic Press, Boston 1998 (p. 157).
[2] J. NEUGEBAUER and C. G. VAN DE WALLE, Phys. Rev. Lett. **75**, 4452 (1995).
[3] W. GÖTZ et al., Appl. Phys. Lett. **69**, 3725 (1996).
B. CLERJAUD et al., Phys. Rev. B **61**, 8238 (2000).
[4] S. LIMPIJUMNONG, J. E. NORTHRUP, and C. G. VAN DE WALLE, to be published.
[5] M. BOCKSTEDTE, A. KLEY, J. NEUGEBAUER, and M. SCHEFFLER, Comp. Phys. Commun. **107**, 187 (1997).
[6] S. G. LOUIE, S. FROYEN, and M. L. COHEN, Phys. Rev. B **26**, 1739 (1982).
[7] J. NEUGEBAUER and C. G. VAN DE WALLE, Mater. Res. Soc. Symp. Proc. **378**, 503 (1995).
[8] J. H. EDGAR, S. STRITE, I. AKASAKI, H. AMANO, and C. WETZEL (Eds.), Properties, Processing and Applications of Gallium Nitride and Related Semiconductors, EMIS Datareviews Series No. 23, The Institution of Electrical Engineers, London 1999.
[9] C. STAMPFL and C. G. VAN DE WALLE, Phys. Rev. B **59**, 5521 (1999).
[10] A. RUBIO et al., Phys. Rev. B **48**, 11810 (1993).
[11] C. G. VAN DE WALLE, S. LIMPIJUMNONG, and J. NEUGEBAUER, Phys. Rev. B **63**, 245205 (2001).
[12] C. STAMPFL et al., Phys. Rev. B **61**, R7846 (2000).